Emotions in medical education: A multi-study examination of relations between appraisals, emotional states, and performance within authentic learning environments

Melissa Catherine Duffy
Ph.D. in Educational Psychology
Department of Educational and Counselling Psychology
McGill University
Montréal, Québec, Canada
May 2016

A thesis submitted to McGill University in partial fulfillment of the requirements of the Degree of Doctor of Philosophy in Educational Psychology

Table of Contents

Abstract

Emotions have been shown to play an important—albeit complex—role in learning and performance within education. However, within the domain of medicine, emotions have been largely overlooked. Recently calls have been made to further examine emotions within medical education. In this dissertation, I first conduct a comprehensive literature review that examines how theories and methods from relevant disciplines can be applied to address the role of emotions in medical education. In this review manuscript, limitations in previous research are identified and suggestions for advancing the field include: examining a range of positive and negative emotions, exploring the role of regulation, and testing the relevance of control-value theory within medical education. Moreover, the review demonstrates a need to examine state emotions for authentic learning tasks. On the basis of this review, two empirical manuscripts are presented that aim to address these gaps. The first empirical manuscript reports on a study that triangulates questionnaires, interviews, observations, and physiological data for an authentic surgical task. In doing so, it is concluded that emotions relate to control appraisals, performance, and task outcomes in ways that are largely consistent with control-value theory and can help to inform educational interventions and instructional design. The second empirical manuscript tests the generalizability of these findings by including a broader range of authentic medical learning environments. This manuscript provides insight into the validity of a self-report measure of emotions and further establishes support for control-value theory. By re-examining assumptions, this manuscript also identifies areas for further inquiry into the nature and role of emotions in medical education. Theoretical contributions, educational implications, and future directions are discussed in each individual manuscript, as well as the closing chapter of this dissertation, which provides a final discussion of the contributions and future directions.

Résumé

Il a été démontré que les émotions jouent un rôle important, bien que complexe, dans l'apprentissage et la performance en éducation. Cependant, dans le domaine de la médecine, on a largement négligé le rôle des émotions. Récemment, on a réclamé qu'on examine de plus près le rôle des émotions au sein de l'éducation médicale. Dans cette thèse, je premier conducteur une revue exhaustive de la litérature qui examine comment les théories et les méthodes provenant de disciplines pertinentes peuvent être appliquées afin d'aborder le rôle des émotions dans la formation médicale. Dans cette manuscrit revue, les limites dans la recherche précédente sont identifiées et des suggestions pour faire avancer le domaine comprennent: l'examen d'une gamme d'émotions positives et négatives, l'exploration du rôle de la réglementation, et finalement, tester la pertinence de la théorie du contrôle-valeur dans l'enseignement médical. En outre, cette revue démontre la nécessité d'examiner les émotions pour des tâches d'apprentissage authentiques. Sur la base de cet examen, deux manuscrits empiriques qui visent à remédier à ces lacunes sont présentés. Le premier manuscrit empirique présente une étude qui comprend une triangulation des questionnaires, interviews, observations et données physiologiques pour une tâche chirurgicale authentique. Ce faisant, on conclut que les émotions se rapportent au fait de contrôler les évaluations, les performances et les résultats de la tâche d'une manière qui sont en grande partie conforme à la théorie de contrôle-valeur. Cela peut aider à éclairer les interventions éducatives et la conception pédagogique. Le deuxième manuscrit empirique teste la généralisation de ces résultats en incluant un plus large éventail d'environnements d'apprentissage médical authentiques. Ce manuscrit donne un aperçu de la validité d'une mesure d'auto-évaluation des émotions et établit un soutien supplémentaire pour la théorie de contrôle-valeur. En réexaminant les hypothèses, ce manuscrit identifie également les domaines pour une nouvelle enquête sur la nature et le rôle des émotions dans la formation médicale. Les apports théoriques et les implications éducatives, ainsi que les orientations futures sont discutés dans chaque manuscrit individuel, ainsi que dans le chapitre qui clôt cette thèse.

Dedication

I dedicate this dissertation to my parents, Jeanette and Kevin Duffy.

Your leadership and resilience has been an inspiration.

Acknowledgements

Firstly, I would like to extend my gratitude to my supervisor, Dr. Susanne Lajoie, for your mentorship, advocacy, and encouragement throughout this process. Your faith in me as a scholar has been an integral part of completing this journey and I grately value the relationship we have developed over the past three years. Thank you to Dr. Krista Muis for your guidance, expertise, and energy. You have mentored me since the beginning of my graduate studies and I am grateful for your continued support and encouragement. Thank you to Dr. Nathan Hall for your thoughtful feedback and insight throughout the dissertation process. The time and attention you have dedicated is grately appreciated. Thank you to Dr. Kevin Lachapelle for providing funding, resources, and expertise to support this research and for connecting me to the broader research community in medical education. Your keen interest in the role of emotions in surgical education has moved this work forward. Thank you also to the members of the Advanced Technologies for Learning in Authentic Settings (ATLAS) laboratory, as well as my co-authors and collaborators. You have helped to strengthen this research by sharing your knowledge, resources, and time.

I would also like to extend a sincere thank you to John Ranellucci, Kyle Hubbard, Marianne Chevrier, Meredith Derian-Toth, Reza Fezi-Behnagh, Tara Tressel, and to all my academic family for cheering me on throughout my graduate studies. The long conversations, brainstorming sessions, and Thomson House visits helped me reach the finish line. A special thank you to Greg for your love, encouragement, and support. Your confidence in me has helped me to keep moving forward.

Finally, thank you to my parents, Kevin and Jeanette Duffy, for your unconditional love and support throughout my educational pursuits. You have provided a remarkable model of perseverance, dedication, and generosity. I work everyday to develop these qualities that you so fully embody.

I also acknowledge the financial support I received during my doctoral studies from the Fonds de Recherché Société et Culture Quebec (FRQSC) and the Learning Environments Across Discplines (LEADS) Partnership Grant of the Social Sciences and Humanities Research Council.

Preface and Contribution of Authors

I am the primary author on each manuscript and am responsible for their content. I wrote Chapters 1 and 5 independently and Dr. Susanne Lajoie provided feedback. I also wrote Chapter 2 independently. The original version of this chapter was prepared in partial fulfillment of my comprehensive exam, which Dr. Lajoie, Dr. Nathan Hall, and Dr. Krista Muis provided feedback on in their roles as members of the comprehensive exam evaluation committee. A modified portion of this comprehensive exam (primarily question 2) was published in a book chapter, which was co-authored with Dr. Lajoie and Dr. Kevin Lachapelle. I wrote the first empirical manuscript presented in this dissertation (Chapter 3) independently and Dr. Lajoie provided feedback. A modified version of this manuscript will be submitted to a peer-reviewed journal for publication, co-authored with Dr. Lajoie, Dr. Reinhard Pekrun, Dr. Marina Ibrahim, and Dr. Lachapelle, who each provided input on various aspects of the research as described in more detail below. I wrote the second empirical manuscript presented in this dissertation (Chapter 4) independently and Dr. Lajoie provided feedback. A modified version of this manuscript will be submitted to a peer-reviewed journal for publication, co-authored with Dr. Lajoie, Dr. Pekrun, and Dr. Lachapelle who each provided input on various aspects of the research as described in more detail below. The contributions made by myself, my co-authors, and colleagues are described below for each manuscript. The conclusions drawn from these chapters are considered original and distinct contributions to knowledge.

Chapter 2

Citation

Duffy, M. C., Lajoie, S., & Lachapelle, K. (2016). Measuring emotions in medical education: Methodological and technological advances within authentic medical learning environments. In S. Bridges, L. K. Chan, & C. E. Hmelo-Silver (Eds.) (pp. 181-213). *Educational technologies in contextual learning: Research in health professions education*. Switzerland: Springer.

Contributions

I conducted the literature review and wrote the manuscript in its entirety. The co-authors provided feedback on the full draft.

Chapter 3

Contributions

I was responsible for conceptualization of research questions, measures, protocol, coding schemes, data collection, data entry and analyses, and wrote the manuscript in its entirety. Dr. Susanne Lajoie provided feedback on the full draft. Dr. Lajoie, Dr. Pekrun, and Dr. Lachapelle provided input on research design/feedback on analyses. Dr. Ibrahim constructed the simulation models and assisted with data collection for the simulation component. Dr. Lachapelle provided access to hospital sites, simulation model materials, and funding for the project. Laura Pipe assisted with data entry, as well as transcribing and coding interviews.

Chapter 4

Contributions

I was responsible for conceptualization of research questions, measures, protocol, coding schemes, data entry and analyses, and wrote the manuscript in its entirety. Dr. Lajoie provided feedback on the full draft. Dr. Lajoie, Dr. Pekrun, and Dr. Lachapelle provided input on research design/feedback on analyses. Amanda Jarrell contributed to data collection and calculation of inter-rater reliability for the sample in Study 1. Mikaela Morton and Laura Pipe assisted with data entry, as well as transcriptions and coding of interviews for this sample. Tara Tressel, Tenzin Doleck, and Maedeh Kazemi contributed to data collection for a subset of the sample in Study 2 (deteriorating patient simulation). I was responsible for data collected in the surgical environment and computer-based learning environment in Study 2 (with assistance from Dr. Lajoie's lab members to collect data in the computer-based learning environment).

List of Tables

List of Figures

Appendices

Chapter 1

Introduction

When considering the domain of educational psychology, what often comes to mind is student learning within the classroom in traditional education systems. However, learning and performance do not occur exclusively within these contexts but also within highly specialized and professional domains, such as medicine, which require additional years of intensive training that often take place outside of the classroom in real-world clinical settings and simulation environments. Although medical education has often favored precision, replicability, and quantifiable metrics (e.g., pre-post gains, randomized controlled trials) over subjective experiences and exploratory research, identifying the mechanisms underlying learning and performance requires attention to the *processes* of learning and performance, rather than focusing exclusively on the outcome. Not only are experimental designs a considerable challenge to implement for real-world medical training, the control researchers secure with this design may come at the cost of relevance (Sullivan, 2011). Accordingly, Regehr (2010) advocates for a shift in medical education from a burden of proof and generalizability toward understanding complexity. In a similar vein, within medicine, attention has typically been devoted to cognitive and metacognitive variables due to the inherent links to knowledge and problem solving. However, as medical practice is largely a human endeavor (Swick, 2000), these processes are likely to be influenced by other psychosocial factors (Cuff & Vanselow, 2004), including emotions (Artino & Durning, 2011; LeBlanc, McConnell, & Monteiro, 2014).

Consider, for instance, a surgeon faced with an unforeseen complication during a high-stakes procedure in the operating room. How might the emotions he or she is experiencing—anxiety about the outcome, confusion over which approach to take—affect his/her decision-making or technical performance? Consider an emergency team working to stabilize a patient with a deteriorating condition. How might feelings of frustration or pride influence leadership or decision-making for medical treatment? Not only do these emotions bear on medical professionals' performance but also the learning and performance of trainees, such as medical students and residents.

To address these questions, in this program of research I draw on theories, methods, and constructs commonly tested within educational psychology. The connecting thread among most formal learning environments is that they involve achievement-oriented activities. Achievement activities refer to tasks with an evaluative component that lead to success or failure outcomes (e.g., performing a surgical procedure). High-stakes achievement activities, in particular, are likely to induce more intense emotional states given that they are often characterized by greater time constraints, cognitive complexity, ill-defined problems, and high social pressure for superior performance. Given that these features are exemplified in medical training, this field is uniquely positioned to offer insight into the role of emotions in learning and performance. Moreover, understanding the role of emotions among medical trainees and professionals can help to enrich instructional and educational efforts designed to promote competency and mastery. This is particularly critical in medicine given the impact of performance on patient outcomes. Errors within these high-stakes tasks can lead to increased hospital stays, medical complications, and fatalities (Baker et al., 2004; Kohn, Corrigan, & Donaldson, 1999; O'Haga, MacKinnon, Persaud, & Etchegary, 2009).

As such, the foundation of this dissertation is the theoretical and empirical investigation into the nature and role of emotions in medical education. Beyond this, three key themes further unify this program of research across the three manuscripts: (1) a focus on states rather than traits; (2) a situated approach within authentic learning environments; and, (3) the use of mixed methods to triangulate data. Regarding the first theme, when measuring emotions and related antecedents and outcomes, the goal of this research was to focus on in-the-moment task specific states (momentary) rather than trait-based tendencies (habitual). This approach was chosen given that perceived habits may not correspond to enacted feelings at a given moment in time and may vary across tasks and contexts (Goetz, Bieg, Lüdtke, Pekrun, & Hall, 2013). Regarding the second theme, effective assessment of states requires authentic learning activities that are embedded within real-world training environments. This also ensures the tasks are relevant by heightening ecological validity. Data collected for real-case clinical training is exceptionally rare. This was accomplished by selecting problem-based learning activities that involve active, participatory learning during real-world practice or close approximations (i.e., simulation and clinical training as opposed to classroom or textbook learning)[1]. Finally, the third theme represents a concerted effort to employ multiple data channels to provide complementary information (e.g., interviews, questionnaires, observation/video, physiological data), as well as an effort to incorporate both quantitative and qualitative research traditions to statistically test assumptions and to explore new themes that emerged through participants' perspective.

[1]Although simulation training can take place within the classroom, I refer to classroom here in terms of the more traditional didactic instructional style, which is common in large medical school classes prior to clinical placements and simulation training.

The manuscripts in this dissertation address several complementary research questions as follows: (1) What are the affordances and limitations of various theories and methods to examine and measure emotions (Chapter 2: Literature Review/Manuscript 1)? (2) How do emotions relate to appraisals, regulatory processes, performance, and task outcomes within surgical training in the operating room (Chapter 3: Empirical Study/Manuscript 2)? (3) Is a self-report scale a valid measure of emotions and does control-value theory generalize across authentic medical learning environments (Chapter 4: Empirical Study/Manuscript 3)? In addressing these questions, this dissertation uncovers specific relations between emotions, learning, and performance that have implications for instructional design and educational interventions in medical education, including simulation training.

Overview of the Chapters

Chapter 2 provides a comprehensive review of the literature on theories and methods of conceptualizing and measuring emotions. Here, I identify advantages and limitations of different approaches and their utility within medical education. This chapter also includes a review of relevant empirical studies within medical education and the health professions that have examined the role of emotions and generates novel research questions yet to be addressed.

Chapter 3 presents an empirical study that employed questionnaires, interviews, physiological data, and observations to examine relations between appraisals, emotions, regulatory processes, performance, and learning outcomes for a surgical procedure within a real-case condition and simulation condition within the operating room.

Chapter 4 presents an empirical investigation that builds off the findings from chapter 3 to test that replicability of the relations between appraisals, emotions, and performance across a broader scope of authentic learning environments representing three domains of skills: surgery, diagnostic reasoning, emergency care. The second goal of this study was to examine more closely the validity of the scale used in the previous study, which was specifically designed to measure the unique taxonomy of emotions activated within medical education.

Chapter 5 concludes with a summary of the research presented in this dissertation and its contributions to the advancement of knowledge along with a discussion of the limitations and future directions for research.

References

Artino, A. J., & Durning, S. J. (2011). It's time to explore the role of emotion in medical students' learning. *Academic Medicine, 86*, 275–276.

Baker, G., Norton, P., Flintoff, V., Blais, R., Brown, A., Cox, J., … Tamblyn, R. (2004). The Canadian adverse events study: The incidence of adverse events among hospital patients in Canada. *Canadian Medical Association Journal, 170,* 1678–1686.

Cuff, P. A., & Vanselow, N. (Eds). (2004) *Improving medical education: Enhancing the behavioral and social science content of medical school curricula*. Washington DC: National Academies Press.

Goetz, T., Bieg, M., Lüdtke, O., Pekrun, R., & Hall, N. C. (2013). Do girls really experience more anxiety in mathematics? *Psychological Science, 24,* 2079–2087.

Kohn, L., Corrigan, J., & Donaldson, M. (Eds). *To err is human: Building a safer health system*. Washington DC: National Academy Press.

LeBlanc, V. R., McConnell, M. M., & Monteiro, S. D. (2015). Predictable chaos: A review of the effects of emotions on attention, memory and decision-making. *Advances in Health Sciences Education, 20*, 265–282.

O'Hagan, J., MacKinnon, N. J., Persaud, D., & Etchegary, H. (2009). Self-reported medical errors in seven countries: Implications for Canada. *Healthcare Quarterly, 12,* 55–61.

Regehr, G. (2010). It’s not rocket science: Rethinking our metaphors for research in health professions education. *Medical Education, 44,* 31–39.

Sullivan, G. M. (2011). Getting off the “gold standard”: Randomized controlled trials and education research. *Journal of Graduate Medical Education, 3,* 285–289.

Swick, H. M. (2000). Toward a normative definition of medical professionalism. *Academic Medicine, 75*, 612–616.

Chapter 2: Literature Review/Manuscript 1

Examining emotions in medical education: Theoretical and methodological considerations

A modified version of this chapter (primarily question 2) was published as a book chapter:

Duffy, M. C., Lajoie, S., & Lachapelle, K. (2016). Measuring emotions in medical education: Methodological and technological advances within authentic medical learning environments. In S. Bridges, L. K. Chan, & C. E. Hmelo-Silver (Eds.) *Educational technologies in contextual learning: Research in health professions education* (pp. 181-213). Switzerland: Springer.

Abstract

Emotions serve an important role in learning and performance, yet this construct has been largely overlooked within medical education. In this paper I review theories and methods from relevant fields (e.g., psychology, education, affective sciences) used to detect and understand the nature, structure, and function of emotions. For each major theory and method, I provide a summary of the purpose, identify advantages and disadvantages, and discuss implications for examining emotions within medical education. The goal is to highlight conceptual and methodological considerations. The first section provides a critical review of relevant theories; the second section provides a critical review of relevant methodologies. The conclusion includes a discussion of alignment between theoretical frameworks and methodologies. One outcome of this review is that the ideal approach may be to draw on elements from multiple theories and methods to address distinct, yet complementary, lines of inquiry relating to emotions in medical education.

Keywords: emotion theory; emotion methodology; medical education; literature review

Introduction

Medical professionals are required to make critical decisions that directly impact patient safety and long-term health outcomes. Adverse medical events affect nearly 1 out of 10 patients (de Vries, Ramratten, Smorenburg, Gourma, & Boermeester, 2008) and up to 98,000 deaths per year are attributed to medical errors in the United States alone (Kohn, Corrigan, & Donaldson 2000). Within medical education, emotions may hinder or enhance learning and performance. For instance, positive emotions (e.g., enjoyment) have been linked to more creative and flexible modes of thinking, such as elaboration, critical thinking, and metacognitive monitoring, whereas negative emotions (e.g., anxiety) have been related to more rigid thinking, such as the use of rehearsal strategies (Pekrun, Goetz, & Titz, 2002; Pekrun, Goetz, Frenzel, Barchfield & Perry; 2011). Although researchers have examined the impact of cognitive errors (Bond et al., 2004; Croskerry, 2003) and communication difficulties within medical settings (Leonard, Graham & Bonacum, 2004; Lingard et al., 2008; Manser & Foster; 2011), far less attention has been devoted to the role of emotions for learning and achievement (Artino, 2013; Artino et al., 2012). Recently, calls have been made to address this gap (e.g., McConnell and Eva, 2012; O'Callaghan, 2011; Shapiro, 2011). As Croskerry Abbass, and Wu (2008) argue, there is a need for medical educators to better understand the role of emotions and apply this knowledge to clinical training given the potential impact on decision-making and affect-based errors. Similarly, LeBlanc, McConnell and Monteiro (2014) note that, "…health professions education could be enriched by greater understanding of how these emotions can shape cognitive processes in increasingly predictable ways" (p. 1).

Within medicine and health sciences, emotions have typically been assessed in terms of psychological health, burn-out and well-being (e.g., Chew et al., 2013; Dyrbye et al., 2014; Satterfield & Hughes, 2007) rather than their functional role in learning and performance. Research that has examined links between affect-related variables and clinical performance have focused primarily on the role of stress (e.g., Harvey, Bandiera, Nathens, & LeBlanc, 2012; LeBlanc, 2009; LeBlanc et al., 2011; Piquette et al., 2014), which is typically considered to be a negative state that aligns closely with anxiety (Sharma & Gedeon, 2012). Although this research has contributed greatly to developing a better understanding of the impact of stress, moving forward there is a need to complement this approach by considering a broader range of emotions. Table 1 includes representative examples of studies and measures resulting from a literature search I conducted on research examining the role of emotions within medical and health sciences education. These studies range from emergency medicine and technical skills training to clinical reasoning and communication skills training.

As demonstrated in Table 1, previous research has typically employed a combination of self-report and physiological measures (cortisol levels and heart rate) to identify stress patterns in particular. The few studies within medical education that have examined a broader range of emotions from a learning and achievement perspective have typically relied on self-report measures (e.g., Artino, Rochelle, & Durning, 2010; Fraser et al., 2012; 2014; Hunziker et al., 2011; Kasman, Fyer-Edwards, & Braddock, 2003). In terms of the theoretical framing, emotion theories are often not explicitly stated in these studies. As demonstrated in Table 1, the most frequently cited framework is Lazarus' cognitive appraisal theory of stress (Lazarus, 1966; Lazarus & Folkman, 1984), which focuses on one type of affective phenomenon (stress responses) rather than capturing a range of positive and negative emotional states. As such, in order to effectively address calls to examine emotions within medical education, it is necessary to consider a more diverse array of theories and methods.

With this in mind, in this paper I review theories and methods from relevant fields (e.g., psychology, education, affective sciences) used to detect and understand the nature, structure, and function of emotions. These theories and methods have been selected and reviewed in consideration of their utility for examining emotions within medical education. The goal is to highlight conceptual and methodological considerations. The first question will provide a critical review of relevant theories; the second question will provide a critical review of relevant methodologies. The conclusion will include a discussion of alignment between theoretical frameworks and methodologies.

Scope

In terms of scope, I selected two prominent theories of emotions (i.e., basic and appraisal theories), and two theories that situate emotions within learning and achievement (i.e., control-value theory and emotion regulation). The review addresses key themes for each family of theories rather than describing each individual model. In this review, I opted to exclude theories that did not consider emotion to be unique from other psychological phenomenon or affective states (e.g., psychological construction models [Barrett, 2009; Russell, 2003; Schachter & Singer, 1962]). I also excluded theories that did not focus on emotion as a phenomenon primarily originating within the individual (e.g., social constructionist theories, Mesquita, 2010; 2014; Solomon, 2003), although these alternative perspectives are considered in the broader discussion of the conclusion. Other programs of research that focus exclusively on one type of affective phenomenon (e.g., stress responses, threat appraisals [e.g., Lazarus, 1966]) were not included for review. I discuss the concept of core affect (valence and activation [Russell & Barrett, 1999; Russell, 2003; Yik, Russell, & Steiger, 2011]) in terms of organizing and classifying emotions in the methodology section, rather than as a separate theory as it has been integrated within several frameworks. Given the proliferation of measures, the methodology section focuses on commonly used data channels and methods of classifying emotions. Furthermore, with the exception of the section directly addressing emotion *regulation*, the majority of this review focuses on emotion *generation*. In terms of application for medical education, this review focuses on authentic practical training for medical trainees (i.e., activities that proximate or simulate the types of tasks and skills required for medical practice), such as simulation environments, computer-based learning environments, and training within the operating room.[2] Tables 1-3 present summaries of empirical articles, theories, and methods discussed within this paper.

Question 1. What theories can be used to examine emotions within medical education?

Defining Emotions

Emotions are assumed to serve important functions for survival by allowing individuals to react quickly, intensely and in an adaptive fashion (Plutchik, 1980). Emotions are also important for psychological health and well-being and can impact the surrounding social climate (Izard, 2010). But what is meant by the term *emotion*? Although a definition of emotion has not attained broad consensus in the affective sciences (Izard, 2010), some scholars have noted that the use of the word in daily language can provide a helpful foundation to frame this construct (Mulligan & Scherer, 2012). According to Merriam-Webster's dictionary emotion is defined as: "a conscious mental reaction (as anger or fear) subjectively experienced as strong feeling usually directed toward a specific object and typically accompanied by physiological and behavioral changes in the body" (emotion, 2015). Within the research community, emotions have been treated as subjectively experienced feelings (Shuman & Scherer, 2014), which are mentally represented in the mind (Pekrun & Bühner, 2014). Emotions are considered to fall within the

[2] The terms *real-world* and *naturalist*ic are used in this review to refer to training environments that involve real patients (as opposed to mannequins or simulated/hypothetical cases) and are situated within a hospital or medical setting that is recognized and accredited by a governing body.

broader concept of affect, but are distinct from other affective phenomena, such as moods, in that they are more intense, have a clearer object-focus, have a more salient cause (linked to an event and situation), and are experienced for a shorter duration (Scherer, 2005; Shuman & Scherer, 2014). A rich history of scholarly interest in emotions (e.g., Darwin, 1872; James, 1890) has produced a number of theories of emotion. I discuss prominent theories below. Table 2 contains a summary of each theory based on core features, advantages, limitations, and applications for medical education.

Basic Theories of Emotions

According to most basic theories of emotions (e.g., Ekman, 1992; Ekman & Cordar, 2011; Izard, 1993; 2007; 2011; Plutchik, 1980; 2001), emotions serve an evolutionary purpose and as such have a distinct biological basis (Scarantino & Griffiths, 2011). These theorists argue that there is strong evidence to show that the most primitive, elemental emotions (i.e., *basic*) are universal among humans and in many cases present in other primates. According to this theory, these basic emotions evolved, through selection, given their adaptive value in tasks important for survival, such as avoiding predators, fighting, and securing mates (Ekman, 1999). Support for this perspective is largely based on Paul Ekman's seminal research demonstrating consistency in emotion-specific facial expressions and judgments linking these expressions to specific emotions across culture. Although there is disagreement regarding which emotions in particular qualify as basic (Scarantino & Griffiths, 2011), commonly agreed-upon basic emotions typically include some variation of: anger, fear, happiness, sadness, surprise, and disgust (Ekman, 1999; Ekman & Cordaro, 2011). Basic emotion theorists typically view these emotions as categorically discrete from one another in that they are each distinctly linked to: (1) a specific neural network; (2) a fixed set of bodily expressions; and (3) a specific evolutionary function (Gross & Barrett, 2011; Tracy, 2014a; Tracy & Randles, 2011). Given the emphasis on evolutionary and biological basis of emotions, in their purest form, these primitive emotions are considered to be a product of nature (i.e., innate) rather than nurture (i.e., learned) and involve an automated set of responses.

While basic emotion theorists acknowledge that socialization and enculturation processes can give rise to variations in emotion expression and a more complex range of higher-order emotional experiences during adulthood, they argue that these emotion variants or blends of emotions (Plutchik, 1980, 2001) can be traced back to the original family of basic emotions and the respective subcortical brain structures (Tracy & Randles, 2011). For example, frustration and irritation would be considered members of the anger family. Critiques of basic emotion theorists have often contended that this strong evolutionary stance downplays the role of social factors and experience in emotion generation (Scarantino & Griffiths, 2011). Other issues include the lack of a robust criteria (necessary and sufficient conditions) to determine exactly what makes an emotion basic (Ortony & Turner, 1990) and the lack of a universally accepted taxonomy or list of true basic emotions (Nesse & Ellsworth, 2009). For example, does an emotion need to have a distinctive facial expression to qualify as basic? Does it need to be universal? Does it need to be present in other primates? Or is the true marker of a basic emotion that it served a distinct evolutionary function because it "solved an adaptive problem" (Buss, 2014, p. 313)? Which emotions should be included on the list of basic emotions? Which labels should be used to describe these states?

Setting these disputes aside, basic emotion theory has several common features that are useful to consider when examining emotion generation within medical education settings, particularly within crisis or high-stakes situations (e.g., emergency rooms, surgical settings) where basic emotions are arguably more likely to be activated (Tracy & Randles, 2011). For instance, the principle that these emotions evolved over time due to their utility for reproduction suggests that they are not necessarily designed to be adaptive within medical situations (e.g., sustained fear in response to a threat) depending on the goals and context. Determining whether or not these emotions serve an adaptive function within medical education settings requires empirical scrutiny. The tenants of basic emotion theory also suggest that basic emotions may be less susceptible to impression management tactics (attempts to conceal emotion expression) given their innate and automated nature among humans; in other words, these basic emotions may be more consistently expressed and detected across medical education environments when activated. Given that these emotions are more automatically activated, they may require regulation strategies that involve monitoring and managing them rather than preventing their occurrence. Acknowledging and normalizing their existence may be an important factor in effective regulation within medical education. In accordance with a basic emotion view of these emotions as primitive and thus more highly activated during earlier stages of development (Tracy & Randles, 2011), it may be the case that these emotions are more intensely activated among novice medical trainees compared to experts. Following this premise, the medical socialization process may help to curb the expressions or experiences of basic emotions.

Despite the utility of basic emotion theory in understanding the more instinctual and universal emotional responses of humans, emotion activation is largely treated as an innate and automated process within this framework, which gives little consideration for the role of higher-order cognitive processes, such as appraisals, in the activation of emotion. Rather than vaguely acknowledging the role of perception or meaning making, appraisal theories place subjective cognitive perceptions as an essential component of emotion generation and present new conceptualizations regarding the nature of emotions; these theories are discussed next.

Appraisal Theories of Emotions

According to most appraisal theories of emotions (e.g., Ellsworth, 2013; Frijda, 1986; 2007; Lazarus, 1991; Scherer, 1984; 2013; Smith & Ellsworth, 1985), subjective cognitive appraisals or perceptions lie at the heart of emotion generation. Appraisal theories treat emotions as multi-componential, in that each emotion is composed of coordinated sets of psychological sub-systems, which, for most models, include cognitive, motivational, expressive, and physiological components, as well as a subjective feeling (awareness of change in one or multiple other components [Ellsworth, 2013; Moors, 2014; Moors, Ellsworth, Scherer, & Frijda, 2013]). This represents a departure from basic emotion theories in that emotions are not conceptualized as a finite set of discrete categories linked to designated brain mechanisms, but rather emotions are considered to be multi-componential processes or episodes that are continuous and recursive (Scherer, 2005; 2009; Shuman & Scherer, 2014). At any point in time these components are simultaneously activated to varying degrees and collectively produce an infinite number of emotional experiences, which may or may not be aligned with a recognized emotion label or word (Ellsworth, 2013, Moors et al., 2013).

According to appraisal theories of emotions, an individual detects and assesses an event or situation based on any number of appraisal variables (e.g., perceptions of control, goal relevance, goal congruence, urgency, certainty, agency, novelty), which in turn can influence the intensity and quality of other sub-component processes of emotion (e.g., facial expression, physiological arousal) (Moors et al., 2013). The recursive nature also allows these sub-components to influence appraisals (e.g., excessive sweating may lead you to believe you have lost control over the task at hand) but ultimately it is appraisals that are responsible for differentiating emotions (Moors et al., 2013). These appraisals are often unconscious or quick but can also be less automated (Moors et al., 2013) particularly when an event is perceived as novel (Ellsworth, 2013). The subjective and changing nature of many possible appraisals values and combinations (e.g., high/low control + high/low goal congruence) allows for variability in the emotional experience within and across individuals and environments and provides a contextually-sensitive framework (Gross & Barrett, 2011). That is, this theory explains how individuals may experience vastly different emotions in response to seemingly similar events given that their appraisals may be different (Ellsworth, 2013).

Although it is often assumed that appraisals are separate antecedents that *cause* emotions, appraisal theorists typically view appraisals as a key component or ingredient that influences other components of emotions (e.g., motivation) rather than a pre-emotional state (e.g., Ellsworth, 2013). In this sense, it is unclear whether appraisals are separate antecedents to emotions or whether they are themselves a component. Another issue is the lack of agreement about the number, nature, and relative importance of specific appraisal variables (Moors et al., 2013). For instance, how many appraisals variables should be taken into consideration to predict emotions? Which are most important for emotion generation in educational and achievement settings? Should appraisal values be categorical (goal congruence/goal incongruence) or continuous (low to high goal congruence)? Furthermore, the mechanism and automaticity underlying the influence of appraisals on other components of emotions is not well established or articulated (Ellsworth, 2013; Moors, 2014; Moors et al., 2013). For instance, do appraisals influence each component at the same time? Do they influence the motivational component primarily, which in turn influences the remaining components? How automated is this process? Under what conditions is it more self-conscious and deliberate versus unconscious and habitual?

Despite these unresolved debates, appraisal theories have several implications for examining emotions within medical education. First, these theories allow for a wider range of emotions than basic emotion theory. A rich array of nuanced emotional experiences is likely to occur within medical training environments given the multifaceted nature of this culture (e.g., hierarchical structure, complex tasks and decision-making, social nature of teamwork and evaluations, high performance demands, high-stakes situations). Secondly, invoking appraisals may be particularly useful for medical education as differences in trainees' appraisals may account for emotion variations across learning environments and levels of experience. For instance, during simulation training, novices may feel greater control over the task at hand compared to real-world training settings and consider it to be highly congruent to their learning goals (moderate challenge and opportunity to improve skills), which may produce positive emotions (e.g., enjoyment), whereas experts may also experience high control but relative to real-world training perceive simulation training as incongruent to their learning goals (low challenge, limited opportunity to improve skills), which may produce negative emotions (e.g., anger).

Another relevant feature of appraisal theories for medical education is the conceptualization of emotions as multi-componential. Using anxiety as an example, a medical trainee may feel anxious as demonstrated by: (1) ruminating thoughts about performing poorly (cognitive), (2) a desire to leave a situation to avoid demonstrating incompetence among peers (motivational), (3) a feeling of unease and displeasure (subjective conscious feeling), (4) displaying a frown (expressive), and (5) increased sweating (physiological). This suggests that the emotion can be displayed or detected through multiple components, which is particularly important within medical environments where emotions may be less overtly expressed among trainees concerned with displaying confidence in front of staff or instructors. Bearing in mind the recursive nature of emotions, it also suggests that changes in one component (e.g., thoughts, actions, expressions) may influence other channels, in turn changing the emotion experienced. For instance, if a medical trainees change their posture to a more open stance, a shift in emotion toward a more positive state may follow.

Although general appraisals theories provide a fruitful conceptualization of emotions, it is also worthwhile to consider context-specific appraisal models that move beyond the underlying *structure* and *nature* of emotions, toward understanding its *functional role* in the context of education. Control-value theory offers a framework for examining the activation and influence of emotions within the context of learning and achievement. This theory is discussed next.

Control-Value Theory of Achievement Emotions

Control-value theory of achievement emotions (Pekrun, 2006; Pekrun, Frenzel, Goetz, & Perry, 2007; Pekrun & Perry; 2014;) is an appraisal theory of emotion but focuses specifically on the role of emotions within achievement contexts, which according to Pekrun and Linnenbrink (2014) can be defined as: “activities or outcomes that are judged according to competence-related standards of quality” (p. 260). Thus, factors important for academic learning and achievement are taken into consideration within this framework. Furthermore, unlike other appraisal theories, the taxonomy of emotions and appraisals included in this framework were selected specifically for their relevance for achievement activities and outcomes (e.g., enjoyment, hope, pride, relaxation, relief, gratitude, anger, frustration, anxiety, shame, boredom, sadness, disappointment, and hopelessness [Pekrun, 2006, Pekrun & Perry, 2014]). These achievement emotions are classified or organized according to three dimensions: valence (positive versus negative), arousal (activating versus deactivating), and object-focus (activity-related versus outcome-related)[3]. Control and value appraisals were selected for this theory because they are considered to play a critical role in predicting emotions within achievement contexts, as the name of the theory implies, and as a growing body of research supports (see Pekrun & Perry, 2014).

As described by Pekrun (2006; 2007; 2014), *control* appraisals are complex constructs, involving self-concepts of ability and self-efficacy, expectancies related to actions and outcomes, and attributions for success and failure. These control appraisals may be activated prospectively (in anticipation of success or failure outcome) or retrospectively (once success or failure outcome has occurred). To elaborate, prospective control appraisals refer to an individual's perception of control over the successful execution of their actions (action-control expectancy), the perception of whether these actions (e.g., investing effort) will produce a positive outcome (action-outcome expectancy), as well as the perception of whether the outcome would occur even without the individual's actions (situation-outcome expectancy). Collectively, these control appraisals shape the perceived likelihood that a desired outcome will occur. Retrospective control appraisals involve perceptions of whether the outcome was controllable by oneself or external factors (locus of control attributions for success or failure). *Value* appraisals involve perceptions of the intrinsic or inherent value of an activity (e.g., task is perceived as interesting) and the extrinsic or utility value of an activity (e.g., useful for achieving long-term goal) (Pekrun, 2006; Pekrun et al., 2007; Pekrun & Perry, 2014). Value can also refer to the valence (Frenzel, 2013) of an activity (e.g., pleasant versus unpleasant) and outcome (success versus failure).

Different combinations of control and value appraisals are expected to predict specific patterns of achievement emotion activation. For example, in terms of prospective outcome emotions (emotion directed toward possible outcome), if an individual perceives high control and focuses on success, they are expected to experience anticipatory joy, whereas if an individual perceives only partial control and focuses on failure, they are expected to experience anxiety.

[3] These methods of classifying emotions are discussed further in the methods section of this review paper.

With respect to activity emotion (emotion directed toward task at hand), if an activity is perceived to be controllable and is positively valued, then an individual is expected to experience enjoyment (Pekrun, 2006; Pekrun et al., 2007), whereas, if an activity is perceived to be controllable but is negatively valued, then anger is likely to occur (Pekrun, 2006; Pekrun et al., 2007). With respect to retrospective outcome emotions, if the outcome is deemed successful and perceived to be caused by oneself (e.g., effort/ability), then pride is expected, whereas if the outcome is deemed a failure and caused by oneself (e.g., effort/ability), then shame is expected[4].

According to control-value theory, appraisals are influenced by a host of factors (distal antecedents) such as individual differences (e.g., individuals' beliefs, gender, and achievement goals) and broader aspects of the learning environment (e.g., task demands, goal structures, epistemic climate, autonomy support [Pekrun, 2014). For instance, if the instructional strategies in the classroom convey messages about the goal structure that align with the learner's personal achievement goals, this may promote perceptions of greater personal control and value for the learning task, which in turn may promote more positive emotions, such as enjoyment (Pekrun, 2006; Pekrun et al., 2007; Pekrun & Perry, 2014). Once activated, emotions are expected to influence performance through the mediated role of cognition (e.g., working memory, information processing), motivation (e.g., persistence, effort), and self-regulated learning (e.g., monitoring understanding, evaluating progress) (Pekrun et al., 2007; Pekrun & Perry, 2014).

As an appraisal framework, control-value theory is susceptible to the same limitations as general appraisal theories and also has shortcomings of its own. For example, although this theory offers a much-needed focus on achievement emotions, it overlooks other academic emotions, such as social emotions (e.g., gratitude, compassion, jealousy), that may be relevant for achievement settings. In addition, there is some risk of over-simplification due to the dichotomization of appraisal values (e.g., high versus low control) and outcomes (e.g., success versus failure). As in any case of transforming a potentially continuous variable into a categorical one, some information is lost. It is likely that other relevant learning outcomes besides achievement, such as resilience, intrinsic motivation, and conceptual change are more appropriately conceptualized as continuous. Furthermore, it is not clear how these control and value combinations account for different sub-types of a particular emotion. For example, recent

---

[4] For the complete list of predicted patterns see Pekrun (2006, 2014). It is not provided here as it is beyond the scope of the review.

research (e.g., Goetz et al., 2014) has found evidence for different types of boredom that vary in valence (positive versus negative) and arousal level (activating versus deactivating). Can control and value alone account for nuanced emotional experiences? Should other appraisal variables be considered? Another consideration is that although control and value appraisals are presented as multi-faceted constructs, many studies have opted to report a global measure of controllability or value rather than reporting the sub-components separately (e.g., intrinsic and extrinsic value; action-expectancies and action-outcome expectancies). As such, the relative importance of these sub-components of control and value variables is not evident.

Although control-value theory provides reasoning for the classification of emotions as activity versus outcome-related, these types of assumptions need to be further tested. For example, pride is classified as a retrospective outcome emotion (occurs after success) but it could also be classified as a prospective outcome emotion if an individual views the upcoming activity as an opportunity to display their competence and anticipates success. Lastly, this theory does not fully elucidate the moment-to-moment changes in emotional experience from a process perspective (i.e., temporal dynamics). For example, what happens when emotions change in intensity over time or multiple emotions co-occur at a given point in time? During which key events are outcome versus activity emotions activated?

Despite these remaining questions, control-value theory arguably offers the most relevant framework to examine activity and outcome-related emotions within medical education (Artino, 2012). For instance, medical trainees may experience enjoyment during an interesting diagnostic reasoning task, boredom during a dull lecture, or pride upon receiving positive clinical skills assessment from an instructor (Artino, 2012). In terms of empirical research, there is some evidence to suggest that within medical classrooms, positive emotions, such as enjoyment, may be more beneficial for achievement than negative emotions, such as boredom or anxiety (Artino La Rochelle, & Durning, 2010). Research within medical simulation environments indicates that positive activating emotions in particular (e.g., invigoration) are positively related to performance whereas positive deactivating emotions (e.g., tranquility) are negatively related (Fraser et al., 2012). Consistent with these patterns, negative emotional states have been linked to lower-order cognitive and metacognitive states during simulated medical emergencies (Duffy et al., 2015a).

However, as the broader emotion literature demonstrates, both positive and negative emotions can focus attention away from the task and lead to task-irrelevant thinking by redirecting attention toward the object of the emotion (Meindhardt & Pekrun, 2003; Pekrun et al., 2007). This depends to some extent on what cognitive processes and behaviors are motivated by the emotion. For example, anxiety can be useful if it leads to increases in effort but ineffective if it interferes with attention (Pekrun, 2006; Pekrun et al., 2007). This is consistent with research in medical education that has demonstrated complex relationships between anxiety and performance; in some cases, anxiety is positively related to performance (e.g., LeBlanc & Bandiera, 2007; LeBlanc et al., 2008), whereas in others it has a negative relationship (e.g., Harvey et al., 2012, Hunziker et al., 2011; LeBlanc et al., 2005, 2012). As Harvey et al., (2012) suggest, these differential effects may be related to the complexity of the task: during less complex tasks (e.g., suturing), anxiety may serve to enhance attention, whereas during more complex tasks (e.g., trauma resuscitation), anxiety may place excessive demands on limited cognitive resources and impede performance.

In a similar vein, other negative states, such as confusion may be beneficial when learning about complex medical concepts given that they may serve as a catalyst for conceptual change (D'Mello, Lehman, Pekrun, & Graesser, 2014; Graesser, D'Mello, & Strain, 2014). However, persistent confusion coupled with low control may require intervention to address disparities in prior knowledge and skill level. Based on their review of research examining links between emotions and cognitive processes, McConnell and Eva (2012) suggested that positive emotions may enhance more global big-picture problem-solving and creative thinking among health professionals (potentially useful for avoiding fixation errors) whereas negative emotions may lead to detail-oriented and analytical thinking (potentially useful for tasks requiring attention to protocol). As these examples illustrate, whether or not a particular emotion is considered *adaptive* or *maladaptive* is context-dependent (Pekrun et al., 2007). Teasing apart relations between appraisals, emotion activation, task complexity/urgency, and skill level will be important step toward making training decisions that optimize learning.

Taken together, the theories introduced up to this point (basic, appraisal, control-value) have largely focused on how and why emotions are generated, rather than how they are monitored and controlled, which is an important component in understanding the functional role of emotions for learning and achievement within medical education. In the following section I discuss how regulation theories can be used to better understand how emotions are managed.

Emotion Regulation Theories

Regulation theories of emotion can be grouped according to two classes: those that describe different types of emotion regulation strategies that may be activated in daily life and those that contextualize emotion regulation as it relates to other self-regulated processes involved in learning and achievement. In this section, I draw on both as they each have features that may be beneficial for examining emotion regulation within medical education.

Emotion regulation has been defined as the process of modulating or modifying one's emotional experience and can take several forms, such as decreasing or increasing the intensity or occurrence of an emotional experience, as well as maintaining or prolonging an emotion (Eisenberg, Fabes, Guthrie, & Reiser, 2000; Gross & Thompson, 2007; Koole, 2009). According to Gross' (1998a, 1998b, 2001) process model[5], emotion regulation is distinct from coping strategies, and other modes of affect regulation, in that it is typically shorter in duration, involves a more specific object-focus and includes both positive *and* negative emotions (Gross & Thompson, 2007). Although goal-directed in nature, regulation can range from a largely unconscious and automated process (i.e., an implicit goal) to a more conscious and deliberate effort (i.e., an explicit goal) (Gross, Sheppes & Urry, 2011; Gross & Thompson, 2007). Regulation may occur at various stages of emotion generation. It may occur early in the process to prevent the onset of an undesired emotion (antecedent-focused), or it may occur later in the process in response to an activated emotion (response-focused) (Gross, 2001; Gross & Barrett, 2011; Gross & John, 2003). In both forms, the goal of emotion regulation is to modify a current or anticipated emotion (i.e., the target is the emotion), whereas the goal of emotion generation is to achieve a desired outcome (i.e., the target is outside the emotion [Gross et al., 2011]).

Emotion regulation can be organized into different families of strategies. For example, Pekrun and Perry (2014) have recently built from Gross's (1998a; 1998b) conceptualizations to re-organize regulation strategies into the following four categories: (1) emotion-oriented, which involves modifying any component of the emotion (e.g., suppressing overt expression of emotion); (2) appraisal-oriented regulation, which involves modifying the antecedent appraisals (e.g., reinterpreting performance as successful); (3) situation-oriented regulation, which involves

[5] This is considered the most widely used model of emotion regulation (Webb, Miles & Sheeran, 2012)

selecting or modifying tasks and environments (e.g., choosing a quiet place to study); (4) competence-oriented regulation, which involves efforts aimed at improving knowledge or skill (e.g., practicing a technical procedure). Within each of these, different strategy sub-types can be further identified. Across empirical studies, fairly consistent patterns have emerged demonstrating that reappraisal strategies in particular are more effective in terms of modulating intended emotional states and improving psychological well-being than suppression strategies (e.g., Aldao & Nolen-Hoeksema, 2010, 2012; Augustine & Hemenover, 2009; Webb, Miles, & Sheeran, 2012), although no strong presumptions have been made about which strategies are universally adaptive, as this can vary depending on the context and goals (Gross et al., 2011; Pekrun & Perry, 2014).

Although these frameworks introduce an important construct, theories of self-regulated learning (SRL, Boekaerts, 2011; Efklides; 2011; Pintrich, 2000, 2004; Winne & Hadwin, 2008; Zimmerman, 2011) may provide a more complete illustration of the temporal nature of emotion regulation (Azevedo et al., 2013). In addition, SRL frameworks raise important considerations regarding the nature of regulation and effective deployment of strategies. For example, most SRL frameworks describe regulation as a process that involves continuous and recursive processes of planning, monitoring, control, and evaluation. In traditional SRL frameworks (e.g., Winne, 2001; Winne & Hadwin 1998; Zimmerman 2000) these processes are commonly discussed in terms of cognition and behavior, but they can also be applied to emotional processes (e.g., Ben- Eliyahu & Linnenbrink, 2013; Boekaerts, 2011; Efklides, 2011; Pintrich, 2004; Zimmerman, 2011). Given that emotion regulation frameworks focus exclusively on the deployment of strategies (i.e., a control mechanism), three key components of SRL could enrich our understanding of emotion regulation, including: (1) monitoring processes, (2) evaluation processes, and (3) goal-setting.

According to SRL theories, monitoring and evaluation processes are enacted in relation to an individual's learning or achievement goals, and as such the subjective evaluations regarding the effectiveness of a strategy is dependent on the personal standards for success. For instance, if a learner's goal is to appear competent and confident to his/her instructor, then the choice to suppress overt expressions of anxiety would be selected, deployed, and evaluated in relation to this goal. As Tamir (2011) has noted, emotion regulation research should investigate both *how* and *why* emotions are regulated. The implications are that emotion regulation should be examined as a continuous process that involves setting goals for ideal emotional states, checking on current and forecasted emotional states to determine if they are aligned with goals, deciding whether a strategy should be used to adapt emotions, selecting an appropriate strategy, evaluating the effectiveness of the strategy in terms of whether it successfully modified the intended emotional state, and in the case that it was not effective, deploying a new strategy or adopting a new goal.

One limitation of emotion regulation frameworks is that the distinction between generation and regulation can become nebulous, as these may be co-occurring episodes (Gross & Barrett, 2011; Shuman & Scherer, 2014). In other words, how does one differentiate the change of emotional state in response to a new event or appraisal from the change in emotional state as a result of regulation? The concept of meta-emotion (i.e., secondary emotions that arise in reaction to a current emotional state, Mitmasgruber, Beck, Höfer, & Schübler, 2009) introduces similar challenges given that the object focus of the emotion becomes emotion itself. For instance, an individual may feel angry if they are making mistakes during a performance task and then subsequently feel embarrassed that they were angry if they think it was an inappropriate reaction within a social setting. In such cases, it is likely that inferences distinguishing emotion-*generation* from emotion-*regulation* will vary in terms of their certainty. To some extent, clarity depends on the theoretical perspective adopted (Gross & Barrett, 2011), as emotion regulation is not anchored exclusively to a single emotion theory (Shuman & Scherer, 2014). In basic emotion models, emotion generation is postulated to reside primarily in subcortical brain structures, whereas regulation is a higher-order control process involving cortical brain structures. However, in appraisal models, some components of emotion-generation (e.g., appraisals) involve cortical brain structures and, therefore, may become more difficult to distinguish from regulation (Gross & Barrett, 2011). However, these links to neural pathways are largely theoretically and have not been rigorously tested within educational psychology at this point.

In addition, other questions concerning the nature of regulation have not been addressed within existing frameworks. For instance, under what conditions is regulation a deliberate versus automated process? Does habituated regulation reflect expertise? To what extent should a regulation strategy focus on emotion as the primary versus secondary object of regulation? To what extent are emotion regulation strategies domain-specific? To what extent does the feeling component of emotion relate to the monitoring component of regulation? A major limitation of existing emotion regulation taxonomies is that they often emphasize modulating negative affective states (e.g., anxiety), rather than positive emotions (e.g., maintaining and increasing enjoyment [Tugade & Fredrickson, 2006]). In addition, these taxonomies tend to favor the control component of regulation (i.e., strategy deployment) and overlook other key regulatory processes (e.g., monitoring current emotional states, evaluating effectiveness of a regulation strategy). Furthermore, considerable variability can be found across studies in the way they have defined, manipulated or measured a particular strategy (Webb et al., 2012). For example, a meta-analysis of emotion regulation strategies (Webb et al., 2012) found seven forms of attention deployment strategies (e.g., active versus passive distraction), four forms of reappraisal strategies (e.g., reinterpreting events versus accepting emotion), and three forms of emotion suppression strategies (e.g., suppressing observable expression versus suppressing experience of emotion). These findings suggest that attention should be paid to the way in which strategies are deployed.

Despite these limitations, emotion regulation has important implications for learning and performance within medical education, as well as interventions designed to help trainees' up-regulate adaptive emotions and down-regulate maladaptive emotions. Given that it accounts for unique variance in performance and well-being (Gross & Barrett, 2011), the inclusion of emotion regulation arguably offers a more complete picture of medical trainees' emotional experiences and goal-striving compared to emotion generation alone.

Within surgical settings, researchers have identified strategies used to deal with negative affective states, including: breathing techniques, distraction, self-talk, help-seeking, team briefing, rumination, denial of responsibility, resignation, and situation avoidance (e.g., Arora, 2009; Hassan et al., 2006; Wetzel et al., 2010). In this context, certain forms of regulation typically considered adaptive (e.g., help-seeking) may be viewed as maladaptive given that they may interfere with the goal of completing an urgent technical procedure (Hassan et al., 2006). Furthermore, while situation-oriented and competence-oriented strategies may be feasible for classroom contexts and studying activities that afford greater autonomy, emotion-oriented and appraisal-oriented strategies may be more adaptive within simulation centers, clinical settings, or operating rooms, given that the urgency of task completion necessitates immediate action.

It may also be the case that certain regulation processes are more accessible or commonly deployed based on level of expertise. Perhaps, for instance, novices more actively monitor their emotions, compared to experts, but lack knowledge or skill to efficiently select and deploy regulation strategies. Perhaps experts regulate in a more automated fashion by preventing an emotion before it is generated, which may help to reduce costs of deliberately attending to and controlling emotions (Mauss, Cook, & Gross, 2007). In any case, regulation seems to be a pertinent skill for medical trainees to master given the potential intensity of their emotions.

Preliminary research (Hunziker et al., 2013) has demonstrated that the use of task-focusing strategies during simulated resuscitation training helps to reduce anxiety among medical trainees but fails to improve performance. However, much of the research in medical education to date focuses on broader constructs of coping rather than emotion regulation and commonly administers scales that are confounded by measures of negative emotions (e.g., worry, anxiety; McWilliams, Cox, & Enns, 2003). By drawing on concepts from self-regulated learning and emotion regulation theories, variability in the prevalence, nature, and effectiveness of different emotion regulation processes can be more closely examined across environments, tasks, and level of expertise within medical education.

As evidenced by my review of emotion theories, several questions regarding the nature, structure, and functions of emotions emerge, each of which has methodological and analytical implications. Despite these unresolved issues, emotion research continues to advance, as do methodological developments. Perhaps, as Picard (2010) and other affective computing (Calvo & D'Mello, 2012) and neuroscience researchers (Immordino-Yang & Christodoulou, 2014) have suggested, data collected from advanced emotion-detection technologies and contemporary methodologies may provide insights that allow researchers to overcome conceptual hurdles and enrich theoretical models. The following section provides a critical review of methods used to examine emotions and their utility for medical education. I start with an overview of how emotions and measures can be organized.

Question 2. What methods can be used to examine emotions within medical education?

Organizing and Classifying Emotions

Emotion measures and their corresponding analytic techniques can be organized based on conceptual features and methodological approaches. For example, from a conceptual standpoint, emotions can be grouped according to valence: positive emotions (pleasant valence) versus negative emotions (unpleasant valence); physiological arousal: activating versus deactivating (Pekrun et al., 2002; Pekrun et al., 2007); and action tendency: whether the emotion is associated with a tendency to approach stimuli (e.g., anger, excitement) or avoid stimuli (e.g., anxiety)

(Frijda, 1986, 1987; Mulligan & Scherer, 2012). Dimensions of valence and arousal are the most commonly used and are consistent with the circumplex model of core affect (Russell, 2003; Russell & Barrett, 1999; Yik, Russell, & Steiger, 2011).

The advantage of using these dimensions (e.g., high arousal, negative valence) to describe emotions is that they provide a more systematic method of differentiating emotional experiences and avoid the issue of relying on folk lexicon used in daily language that may hold different meanings and interpretations across individuals (Yik, Russell, & Steiger, 2011). On the other hand, discrete categories of emotions (e.g., angry, happy, sad) hold valuable meaning for the individual and have the advantage of efficiently capturing distinctions between seemingly similar dimensional states (e.g., anxiety versus excitement), as well as states that appear to lack valence (e.g., surprise). It should be noted that these approaches are not necessarily incompatible (Izard, 2007; Maus & Robinson, 2009; Scherer, 2005; Shuman & Scherer, 2014). In fact, researchers have found evidence for consistency in judgments of associations between dimensions and specific emotion words across samples and languages, which suggest that discrete categories of emotion terms can be grouped into families based on their underlying or higher-order dimensional structure (Fontaine et al., 2013; Shuman & Scherer, 2014; Scherer, Shuman, Fontaine, & Soriano, 2013).

Emotions can also be classified according to the object or event focus, which refers to the focus of the attention (Pekrun & Linnebrink, 2014; Shuman & Scherer, 2014). For example, *outcome emotions* relate to the prospective outcomes (e.g., anticipation of a math test) whereas a*ctivity emotions* relate to the task at hand (e.g., during problem-solving) (Pekrun, 1992; Pekrun et al., 2002). A distinction can also be made based on temporal generality of an emotion: *state emotions* (occur for a specific activity at a specific point of time) or *trait emotions* (recurring emotions that relate to a specific activity or outcome). Emotions can also be organized based on taxonomies, which consist of discrete emotions (e.g., frustration) that are likely to be activated during specific activities or within specific contexts (e.g., academic emotions).

In addition to these conceptual groupings, emotion measures can also be organized based on the methodological approaches employed. For example, emotion measures can be distinguished according to the type of data channel used (e.g., self-report, behavioral physiological), the frequency of data points (e.g., one point in time versus continuous process measures), and the time of administration in relation to the event of interest (e.g., concurrent versus retrospective). Overall, these systems of organizing emotions can be used to identify underlying assumptions, as well as the benefits and limitations of each measure. These features will be taken into consideration in the following review of emotions measures, which are organized according to the type of data channel (self-report, behavioral, physiological). Table 3 contains a summary of each data channel, including example measures, advantages, limitations, and applications for medical education.

Self-Report Measures

Self-report measures of emotions include data channels that rely on individuals' perceptions and communication of emotional states, typically collected through questionnaires and interviews. Given that emotions are considered to be subjective experiences (Shuman & Scherer, 2014) that can be verbally communicated (Pekrun & Linnenbrink-Garcia, 2014), self-reports will likely continue to be widely used (Shuman & Scherer, 2014). Questionnaires are typically designed to measure frequency and intensity of emotions, whereas interviews can be used to further explore participants' perceptions, experiences, and antecedents related to the onset of emotional states. These types of measures vary greatly in terms of breadth and structure, such as single versus multi-item scales, open-ended questions versus structured ratings, state versus trait perspectives, and retrospective versus concurrent reporting (Pekrun & Bühner, 2014). On the surface, self-reports seem to solely assess the subjective feeling component of emotions yet questionnaires and interviews can also be designed to measure other emotion components by designing questions or scale items that include physiological, cognitive, motivational, and expressive responses corresponding to each unique emotion (Pekrun & Bühner, 2014).

Questionnaires can be administered in several manners, including tradition paper-based formats, as well as electronic and web-based questionnaires accessible through computers and Wi-Fi-enabled hand-held mobile devices, such as tablets and smartphones (see Figure 1). These electronic questionnaires afford the flexibility to unobtrusively measure emotions in diverse environments. They also allow responses to be reviewed in real-time, which can be useful for identifying key emotional events to examine further during follow-up interviews or adaptive learning systems. However, there are several unresolved questions regarding the ideal frequency and timing of administration. For instance: how often should a questionnaire be administered to provide a representative picture of concurrent emotions during learning? Should the deployment of questionnaires be aligned to key events or set time intervals? Whereas time-based sampling captures the trajectory of emotions across different phases of learning, event-based sampling can reveal sources of emotions (Turner & Trucano, 2014). Decisions regarding administration will largely depend on the nature of the learning environment and research questions. For instance, researchers may aim to capture salient emotions experienced during learning or they may be interested in linking emotions to specific events, such as the deployment of self-regulated learning processes (e.g., cognitive, metacognitive, motivational) or instructor prompts and feedback. Furthermore, some naturalistic learning environments may require that the questionnaire be administered less frequently to reduce interruptions to the learning activity or performance task at hand.

Experience sampling methods (ESM) (Csikszentmihalyi & Larson, 1987; Hektner, Schmidt & Csikzentmihalyi, 2007) have been used to provide an unobtrusive and representative sample of individuals' emotions in naturalistic settings (e.g., Becker, Goetz, Morger & Ranellucci, 2014; Goetz et al., 2014; Nett, Goetz, & Hall, 2011) by prompting participants to respond to questions at random time points over a specific interval (e.g., over the course of a day, week, or month). This technique has also been used within medical settings to measure trainees' emotional experiences (e.g., Kasman, Fyer-Edwards, & Braddock, 2003). Although this method is intended to capture a representative selection of emotional experiences, it falls short of providing a continuous measure of emotional states at a fine-grained (second-minute) level of temporality. More recently, emote-aloud methods have been used to prompt participants to communicate their emotional states during learning in a manner similar to think-aloud protocols (Craig, D'Mello, Witherspoon & Graesser, 2008); thus, capturing an online trace of emotions and potentially state-transitions as they occur in real-time. However, further work is needed to determine whether this approach is cognitively taxing for participants, whether it is susceptible to social desirability biases or individual variability in verbosity, and how it differs from other speech and paralinguistic channels used to assess emotions (described below).

In terms of the range of emotions detected using self-report measures, the education literature has been largely dominated by a focus on anxiety (Zeidner, 2014). This is also the case for medical and health sciences education as evidenced by the use of the State-Trait Anxiety Inventory (Spielberger et al., 1983) across a number of studies (e.g., Arora et al., 2010; van Dulmen et al., 2007; Harvey et al., 2012; LeBlanc, 2005; Meunier et al., 2013; Piquette et al., 2014). Although these scales have provided important insights into the role of stress in learning, they are complicated by issues with discriminant validity—namely, they appear to assess multiple negative emotions, such as shame and hopelessness rather than providing a pure measure of anxiety, which makes interpretation more difficult (Pekrun & Bühner, 2014). In addition, the exclusive focus on anxiety precludes the study of positive emotions. More recently, the Achievement Emotion Questionnaire (AEQ, Pekrun et al., 2011) has been used to capture a broader range of emotions expected to be elicited within achievement contexts according to the control-value theory of achievement emotions. Several studies (e.g., Goetz, Pekrun, Hall, & Haag, 2006; Pekrun et al., 2007, 2009, 2011; Pekrun et al., 2010) have used the AEQ to assess trait and state emotions (e.g., hope, pride, enjoyment, relief) across various facets of academic achievement settings (e.g., learning and test-taking) at various points in time (e.g., before, during, after the activity). This measure has also been used to assess course-related emotions for medical students (e.g., Artino et al., 2010).

In terms of medical education research, one approach is to use Likert-scale questionnaires and semi-structured interviews to assess state and trait emotions before, during, and after learning and performance activities across authentic learning environments (e.g., simulation training, surgical settings, and computer-based learning environments). Questionnaires can be relatively efficient to administer and analyze across a range of learning environments, whereas interviews can be conducted retrospectively to provide rich contextual information regarding the antecedents, object-focus, regulation, and role of emotions. Collectively, these self-report measures can allow researchers to examine participants' perceptions and subjective experience of emotions using both dimensional features and discrete categories.

Within computer-based medical learning environments (e.g., BioWorld, Lajoie, 2009; Lajoie et al., 2013) time-triggered or event-based questionnaires can be used to measure activity emotions as they occur at multiple points during the session. These concurrent measures are considered to be more valid (Mauss & Robinson, 2009) as they are less susceptible to memory biases and do not appear to interfere with learning within this environment. In contrast, within team-based simulation and surgical settings self-reported activity emotions can be measured retrospectively (via questionnaires and interviews) given that administering questionnaires during the activity may significantly interfere with performance and group dynamics within these achievement-oriented environments.

To conduct follow-up interviews, retrospective and cognitive interviewing techniques can be employed (e.g., Ericsson & Simon, 1993; Muis, Duffy, Trevors, Ranellucci, & Foy, 2014). In addition to gaining information about the nature of emotional experiences, these methods allow researchers to assess convergent validity (i.e., whether there is alignment between interview and questionnaire responses), construct validity (i.e., whether interpretations and responses to questionnaire items align with theoretical assumptions), and face validity (i.e., whether the questionnaire presents a representative range of emotions experienced and is presented in a way that is clear). Eye-tracking and video replay can also be integrated into retrospective interviewing protocols to align participants' recall of emotional states with cued learning events (Van Gog & Jaradzka, 2013).

To collect these types of data, it is important to consider which types of emotions should be included in the self-report measures. Emotion taxonomies can be specifically selected based on their relevance to medical education. For example, within most medical education environments, it is likely that achievement emotions (e.g., anxiety, enjoyment, Pekrun et al., 2011]), social emotions (e.g., compassion, gratitude, Fisher & van Kleef, 2010; Weiner, 2007), and epistemic or knowledge emotions (e.g., confusion, curiosity, Pekrun, Meier, Muis, & Sinatra, under review; Pekrun & Stephens, 2012) are activated given the inherent performance expectations, social dynamics, and knowledge-based nature of medicine. Basic emotions (e.g., fear, anger, sadness, Ekman, 1972, 1992) may also play a role within high-stakes medical education environments, such as simulations and naturalistic settings (e.g., surgical and emergency rooms) given that crisis responses might generate what are considered to be more automated and psychologically primitive or elemental emotions (Tracy & Randles, 2011). This approach is consistent with Shuman and Scherer's (2014) recommendation to include multi-emotion measures, as well as previous work that has adapted self-report measures to include taxonomies of emotions that are more context-sensitive (e.g., Zentner, Grandjean, & Scherer, 2008) given that emotions are considered to be domain-specific and may vary according to the educational environment (D'Mello, 2013; Goetz, Frenzel, Pekrun, Hall, & Lüdtke, 2007; Goetz, Pekrun, Hall, & Haag, 2006) and socio-cultural context (Turner & Trucano, 2014). These developments prompt new questions, such as whether the emotions included are strictly orthogonal (represent independent and unique emotions) or whether they represent several clusters or families of related emotions that share similar components (e.g., sadness and disappointment; Mauss & Robinson, 2009; Shuman & Scherer, 2014).

A key drawback to using self-report measures is that they are susceptible to several validity threats including cognitive and memory biases. Participants may forget the emotional state they are asked to recall or construct a false memory of an emotion to explain their performance. It is also not clear whether participants' interpretation of self-report questions aligns with researchers' assumptions or whether they possess adequate emotional awareness to monitor and effectively describe their emotions. This latter concern may be particularly relevant within medical domains, as it has been suggested that the culture of medicine encourages students to disconnect from their emotions to the extent that they may have difficulty recognizing and processing emotional states (Shapiro, 2011). As previously mentioned, another limitation is that self-reports are not always feasible to administer concurrently during performance tasks within high-stakes environments as the nature of these measures interrupts the task itself. As such, it is worth complementing questionnaire data using less intrusive methods; in the following section, we discuss behavioral measures of emotion.

Behavioral Measures

There is a wide array of behavioral measures, which rely on overt actions, movements, and mannerisms to detect emotional states. Although these measures directly measure the expressive component of emotions, it is argued that these expressions correspond to specific emotional states. Typically, behavioral measures are used to identify the frequency of emotions although intensity can also be assessed depending on the coding methods used (Reisenzein, Junge, Studtmann, & Huber, 2014).

Facial expressions are one of the more commonly used behavioral channels for emotion detection and typically rely on video recordings of facial expressions as they occur and change over time. Coding systems developed to detect these emotional states typically rely on facial movements. While some coding methods (e.g., Kring & Sloan, 2007) focus exclusively on sets of muscle movements that are postulated to represent theoretically meaningful emotions, the Facial Action Coding System (Ekman & Freisen, 1978; Ekman, Freisen, & Hager, 2002) includes a broad range of action units (micro motor muscle movements), and as such, is designed to provide a more objective approach to coding. The FACS approach may also be well suited to natural, subtle expressions given that the facial action unit combinations can result in thousands of possible facial expressions (Zeng, Pantic, Roisman, & Huang, 2009). Automated systems for emotion detection can help to reduce human resources needed to manually code emotions (Azevedo et al., 2013; Zeng et al., 2009). For example, automatic facial recognition software, such as Noldus FaceReader™ (see Figure 2) is designed to classify basic emotions automatically (typically post-hoc) and has been used within a host of studies to detect and trace emotion (e.g., Craig, D'Mello, Witherspoon & Graesser, 2008; D'Mello, Craig & Graesser, 2009; Pantic & Rothkrantz, 2000).

Speech provides another behavioral channel commonly used to detect emotions through discourse patterns and paralinguistics/vocal features (e.g., intonation, pitch, speech rate, amplitude/loudness, see D'Mello & Graesser, 2010, 2012). Previous research has found that combinations of multiple speech features can help to detect specific emotional states (Mauss & Robinson, 2009). Software programs also exist for linguistic and paralinguistic content (Zeng et al., 2009). These programs share similar limitations to facial recognition software in that they are restricted to detecting basic emotions (e.g., happiness, sadness, disgust, fear, anger, surprise), which may not capture the full range or types of emotions activated within educational settings (Zeng et al, 2009). These programs are also based on databases of deliberately displayed (i.e., posed) expressions and, as such, are better equipped to detect extreme or prototypical expressions rather the natural spontaneous expressions we would expect to see in most learning environments (Zeng et al., 2009).

Recent trends have also included analysis of other behavioral features, such as body language and posture (e.g., D'Mello, Dale, & Graesser, 2012; D'Mello & Graesser, 2009). Compared to facial expression and speech features, which appear to be better-equipped to predict valence and arousal dimensions of emotions respectively, specific body postures appear to provide markers for distinct emotional states that cannot be detected as reliably with other behavioral data channels (Mauss & Robinson, 2009). For example, research has shown that there is a link between pride and increased posture, and embarrassment and reduced posture (Keltner & Buswell, 1997; Tracy & Robins, 2004, 2007). An upright posture has also been observed during states of confusion (D'Mello & Graesser, 2010).

In terms of medical education research, facial expression data, body language, and speech/vocal features can be collected during computer-based learning and simulation training using video and audio recordings (e.g., Duffy et al., 2015a). Naturalistic settings (e.g., operating room, clinical setting) may pose practical challenges of installing equipment. As an alternative, field observations can be used to record similar behavioral patterns. Overall, these data channels are relatively inconspicuous and allow participants to complete learning and performance tasks without interruption. They also detect emotions without reliance on participants' knowledge and unbiased communication of their emotional states. Collectively, these behavioral measures provide a concurrent trace that allows researchers to measure the onset and transition of emotional expressions as they occur in real-time. However, the analyses of this data can be time intensive, particularly when human coding methods are employed.

Computer-based learning environments may be ideal for facial expression detection; high quality video recordings can be captured using a mounted camera given that these sessions are typically conducted with one participant at a time and involve relatively stationary behavior. However, this may pose challenges when using a multi-modal approach. For instance, if participants are instructed to sit upright and restrain from covering their face with their hands (to maintain a consistent view of the face), this interferes with the ability to collect recordings of natural body movements, such as leaning towards the computer when curious. Similarly, using think-alouds or emote-alouds for speech and paralinguistic analysis can interfere with the accuracy of facial expression detection as the coding process is reliant on facial muscle movements, including the mouth, which may change as an artifact of speech rather than facial expression.

Simulation environments may allow for more natural body movements and speech behaviors but the issue of multi-modal interference still exists. The increased mobility and dynamic nature of this environment also introduces new challenges, particularly within team-based simulations. For instance, movement can interfere with the ability to consistently capture facial expressions and the distance between the camera(s) and individual team members may not provide ideal conditions to use facial recognition software due to lack of precision and inconsistent facial views. Within these team settings, one option is to track facial expressions, body language, and speech by coding each team member individually and collapsing across the group for analyses and comparative purposes (e.g., team leader versus team members [Duffy et al., 2015a). Emotion events can be segmented according to the onset and endpoint of an emotional cue (as opposed to pre-determined time segments), which is linked to a behavioral channel (e.g., facial expression, body language). Neutral can be included to provide a baseline or resting state (no positive or negative valence) and allow for a continuous stream of emotion codes. This process brings to light several considerations, including: the number of data channels needed to make grounded inference, whether certain data channels are more reliable than others, segmentation of emotional events (time-based versus event-based), the degree of intensity needed to classify an emotion as present versus absent, the co-occurrence of multiple emotions, conflicting evidence from different data channels, and as Järvenoja and Järvelä (2013) have discussed, the occurrence of socially-shared emotional events that appear at the group rather than individual-level.

One advantage of human coding within these environments is that coders can be trained to detect subtle and natural changes in emotion by analyzing micro expressions (Matsumoto & Hwang, 2011) or including contextual cues (e.g., baseline behavior, events preceding emotional event, peer responses, changes over time). This in situ idiographic approach (Picard, 2010) can boost ecological validity by detecting variations that are unique to an individual, group, or task. Another benefit of human coders is that they can employ adapted coding schemes that include a range of emotion taxonomies and corresponding behavioral indices that are context and domain-sensitive. As previously mentioned, most of the automated computer-based coding systems classify basic emotions exclusively. Recent advances in machine learning techniques (e.g., use of log files and eye-tracking for contextual cues) and facial recognition software (e.g., including databases composed of naturalistic facial expressions and learner-centered emotions) may help to address these limitations (Zeng et al., 2009).

Although behavioral measures obviate several limitations of self-report data, they are still influenced by the broader culture of a given domain. As Mauss and Robinson (2009) note, various factors such as social dynamics, feedback, and success/failure perceptions may affect the type and nature of facial expressions displayed, which suggests that expressions may be highly context-sensitive, rather than a one-to-one indicator of specific emotional states (Zeng et al., 2009). For example, in certain situations a smile may be indicative of embarrassment rather than happiness (Mauss & Robinson, 2009). Within medicine, it may be the case that expressions are masked or displayed in a different manner than in other educational environments. In particular, social desirability may reduce emotion expression within medical cultures as participants may consciously or unconsciously conceal observable displays of emotions that they are currently experiencing. Medical students may learn through training and socialization that communicating and displaying one's emotions (whether explicitly or implicitly) is inappropriate (O'Callaghan, 2013; Shapiro, 2011). This may be the reason why behavioral measures of emotions have been less commonly used within medical and health sciences education research, compared to self-report and physiological measures. In the following section, we discuss physiological measures that are less vulnerable to these potential validity threats.

Physiological Measures

Physiological measures of emotion rely on responses of the central and autonomic nervous systems. Central nervous system (CNS) measures attempt to link brain regions to emotions and commonly employ electroencephalogram (EEG) or neuroimaging methods. EEGs record electrical activity to identify which hemispheres (left versus right) and brain regions (e.g., frontal region) are activated during emotional states, whereas neuroimaging technologies (e.g., functional magnetic resonance imaging, positron emission tomography) can identify activation of more precise brain regions (e.g., amygdala) by measuring changes in blood flow and metabolic activity (Mauss & Robinson, 2009). There are several advantages to these types of devices. Similar to behavioral measures, they do not require accurate self-reports or interruption from activities. They also are less susceptible to impression management. These channels provide a fine-grained continuous trace that capture the dynamic nature of physiological arousal as it occurs in real time (Calvo & D'Mello, 2012).

In general, while these measures of the CNS show promise, it is likely that emotional states involve complex networks or systems of brain activity rather than isolated regions; as such, further work is needed to identify circuits of activation (Mauss & Robinson, 2009). At this stage, correlates between brain states and emotions are largely limited to discriminating between positive/negative valence and approach/avoidance action tendency dimensions of emotion rather than discrete emotional states (Mauss & Robinson, 2009) although progress is being made towards linking brain regions to specific emotions, as demonstrated in research linking fear to the amygdala (Murphy et al., 2003). Furthermore, the methods employed to date typically involve provoking a prototypical or extreme emotional state by presenting sensational stimuli in highly controlled experimental settings rather than capturing the more subtle range of naturally-occurring emotions likely to be experienced within academic and achievement settings. Moreover, this methodology is not feasible within applied educational settings at this point (Immordino-Yang & Christodoulou, 2014).

In contrast to the brain regions, peripheral physiological measures of emotions rely on indices of the autonomic nervous system, which is responsible for managing activation and relaxation functions of the human body (Ohman, Hamm, Hugdahl, 2000). Various peripheral physiological measures and affective devices have been used to track emotion-related patterns, including salivary cortisol tests, electrodermal activity/galvanic skin conductance, and heart rate/cardiovascular responses (e.g., Azenberg & Picard, 2014; Matejka et al., 2013; Poh, Swenson, & Picard, 2010; Sharma & Gedeon, 2012). These types of measures, particularly salivary cortisol levels and heart rate, have also been used within medicine and health sciences research to measure stress-related states similar to anxiety during learning and performance (e.g., Arora, 2010; Clarke et al., 2014; van Dulmen et al., 2007; Harvey et al., 2012; Hunziker et al., 2012; Meunier et al., 2013; Piquette et al., 2014). However, given the diverse functions of the autonomic nervous system, it can be difficult to isolate fluctuations that are solely related to emotions rather than other activities of the autonomic nervous systems acting in parallel (e.g., digestion, homeostasis) (Hunzinker et al., 2012; Kreibig & Guido, 2014; Mauss & Robinson, 2009). Moreover, across measures, these channels appear to exclusively measure the physiological component of emotions (high versus low arousal) rather than differentiate between specific emotions (e.g., anger versus fear) (Mauss & Robinson, 2009). Practically speaking, this also makes it difficult to determine whether an increase in physiological activity represents a state of a negative or positive emotion (e.g., anxiety or excitement).

In terms of medical education research, electrodermal activity (EDA) sensors can be used to measure emotion-related physiological arousal before, during, and after learning and performance activities within simulation training, surgical settings, and computer-based learning environments. For example, the Affectiva Q™ sensor bracelet can be used within surgical settings (see Figure 3) as it is portable, non-invasive, and inconspicuous (Poh, Swenson, & Picard, 2010), whereas the Biopac™ sensor can be used within computer-based learning environments (e.g., Dunn, Evans, Makarova, White, & Clark, 2012) given that it requires attachment of electrodes to the skin and a stationary transmitter. In both cases, these devices are arguably better able to isolate emotion-related arousal compared to other physiological measures (e.g., heart rate), which are influenced by non-affective factors to a greater extent (Kreibig & Guido, 2014). These devices can also be used across a range of environments; thus capturing more spontaneous and natural fluctuations compared to brain imaging techniques (Picard, 2010; Poh, Swenson, & Picard, 2010).

These physiological traces can be linked to key events and other data channels (e.g., self-report) by recording time-stamped log files within computer-based learning environments and time-stamped field notes with naturalistic settings (see Figure 4). For example, a novice medical trainee's arousal level may be significantly higher during complex technical procedures compared to more routine procedures and that of an expert. One challenge is that certain EDA devices require a baseline before engaging in the learning activity. Although this is feasible to obtain within lab-based settings, it is more challenging to obtain within naturalistic settings. Furthermore, individuals often possess different baseline measures of arousal/sweat conductance, so exercise or cognitively-taxing tasks may be needed to reach a sufficient threshold for activation (Poh, Swenson, & Picard, 2010). In addition, researchers commonly administer the EDA bracelet on the palmar or forearm, which is not always feasible in clinical settings where hand-washing practices may interfere with device functionality. However, recent research suggests other sites on the body may provide viable options for measurement (Fedor & Picard, 2014).

Another significant challenge is that at this stage discrete emotions are not clearly or consistently differentiated through physiological patterns (Mauss & Robinson, 2009). One approach for analysis is to focus exclusively on reporting levels of arousal rather than emotions. Another approach is to examine how physiological signatures correspond to changes in emotional states using a multi-modal approach (Imordino-Yang & Christodoulou, 2014). For instance, if an individual reports or displays expressions of enjoyment, physiological data can be examined at that precise moment to identify patterns of arousal that may differentiate this emotion from other emotional states using intra and inter-individual analyses. It is also possible that a particular emotion (e.g., boredom) could produce different types of physiological signatures depending on whether an individual is experiencing mixed emotions (e.g., boredom and anger [Larsen & McGraw, 2011]) or different sub-types of a particular emotion (e.g., indifferent boredom versus reactant boredom, Goetz et al., 2014). This type of work may help us to re-examine assumptions about the underlying physiological components of emotions.

At the analytical level, all emotion measures described may involve comparisons across a host of factors, including the following: (1) types of tasks (e.g., diagnostic reasoning, surgical procedures, emergency resuscitation); (1) key events during an activity or task (e.g., novel versus routine procedures) (3) learning environments (e.g., computer-based learning environments, simulation centers, real-world clinical environments, operating rooms); (4) time of response relative to learning or performance (e.g., before, during, after task); and (5) experience level (e.g., novice versus expert). Comparing across these factors may help to illuminate which emotions in particular are more frequently or intensely experienced, which can be used as a index for identifying the most relevant or influential emotions within medical education.

Conclusion

Emotions play an important role in learning and performance, yet this construct has been largely overlooked in medical education. In this paper I have described, analyzed, and evaluated prominent theories and measures of emotions with consideration of their application for medical education. This review illuminates potential avenues of further research to better understand the nature and role of emotions within medical education. The following research questions are guided by the theoretical frameworks and methodologies reviewed: (1) Which emotions (e.g., basic achievement, epistemic, social) are most prevalent or intensely activated in medical education? (2) Do variations in these emotions emerge across environment (e.g., simulation versus clinical settings), task (e.g., diagnostic reasoning versus surgical procedures), experience level (e.g., expert versus novice) and time or event (e.g., before, during, after learning activity)? (4) How do varying degrees of appraisals (e.g., control/value) relate to emotions? (5) How do emotions relate to learning processes (e.g., cognition, attention, regulation) and achievement?

Based on my review, I would argue that the diverse theoretical and methodological perspectives discussed, although each providing unique contributions and foci, are not necessarily incompatible with one another. Despite theorists' attempts to identify discord across perspectives, there is considerable harmony. As Nesse (2014, p. 321) argues: "…a general theory of emotion is neither necessary nor possible." From this vantage point, multiple perspectives provide complementary levels of analysis that can be used to examine emotions (Panksepp & Watt, 2011; Russell, 2014; Tracy, 2014b).

Basic emotion theories suggest that a subset of emotions may be more likely to be intensely activated for high-stakes or novel tasks. This framework is useful for researchers interested in examining the universality of discrete emotions or more primitive emotions at early levels of development (e.g., novice medical trainees). Appraisal theorists argue that subjective appraisals are responsible for differentiating emotional experiences. This framework is useful for researchers examining how medical trainees' perceptions of events influence emotional episodes. Control-value theory situates an appraisal approach within an achievement context by outlining the influence of control and value appraisals on learning processes. This framework is useful for researchers interested in testing relations between medical trainees' emotions, cognitive processes (e.g., attention, memory, problem-solving) and achievement outcomes (e.g., performance on a clinical task). Regulation theories focus on how emotional states are monitored and controlled to achieve a goal. This framework is useful for researchers interested in examining how medical trainees' emotions are monitored and managed.

Certainly, researchers, and even a single study, can espouse more than one goal. Therefore, multiple perspectives will likely improve research design, analyses, and interpretation of results. To illustrate one way that these theories can be brought together: emotional experiences can be assumed to be multi-componential continuous episodes (appraisal theory) but discrete emotional categories can also be applied to clusters or profiles of emotional states that we consider more salient or influential (basic emotion theory). We can situate these emotional states within achievement settings (control-value theory) and examine how they change or sustain over time in response to efforts to achieve a particular goal (regulation theories).

Similarly, measures used to assess emotions are not necessarily in conflict; each focus on a different component of emotions (e.g., subjective feeling, physiological, behavioral). Some channels (e.g., self-report) are more apt to detect discrete states, whereas other channels (e.g., physiological) are better suited to track dimensional categories of valence and arousal. This is consistent with views that dimensional (e.g., valence/arousal) and discrete (e.g., anger, sadness) approaches to measuring and conceptualizing emotions are compatible (Izard, 2007; Mauss & Robinson, 2009; Scherer, 2005; Shuman & Scherer, 2014).

In terms of selecting appropriate measures, it is important to consider practical constraints and limitations inherent to each measure's design. For instance, it may be intrusive to interrupt real-world tasks (e.g., surgical procedures) with self-report measures or cumbersome to install equipment required for detailed facial analyses outside of computer-based learning environments. In addition, self-report measures are efficient and well suited for group comparisons but may be susceptible to cognitive biases, memory errors, and unanticipated interpretations of scale items. On the other hand, behavioral channels can provide an objective measure of emotion expressions but may be time-intensive to analyze if conducted manually. Physiological data channels are not susceptible to attempts to conceal emotions (e.g., social desirability, impression management tactics) and provide a continuous stream of data to allow for analyses of temporal fluctuations; however, at this stage they are less capable of differentiating discrete emotional states (e.g., anxiety versus excitement) and better suited for identifying intensity of arousal levels (high versus low arousal). Given that there is currently no consensus as to the most valid and reliable measure of emotions (D'Mello, 2014; Gross & Barrett, 2011; Mauss & Robinson, 2009), data collection can be maximized by collecting physiological and observational data throughout a learning task and reserving self-report measures for critical times or events (e.g., before, during, after learning). Moreover, by using multiple measures, researchers can examine convergence across channels and more rigorously test assumptions about the underlying structure of emotions.

It should be noted that a particular measure is not necessarily tied to a specific theory. For example, LeDoux (2014) describes that as a neuroscientist he uses brain imaging to measure emotional states but aligns more closely with an appraisal perspective, although his methodological approach is assumed to be as consistent with basic emotion theory. Instead, continuity with a given theory likely depends on which variables are included, how data is collected, and what methods are used to analyze data. Admittedly, challenges surrounding data alignment and accuracy of inferences continue to exist. Furthermore, the measures and theories discussed in this review focus on the individual unit of analyses, whereas emotion generation and regulation may take place at the group level in response to social dynamics or the broader environment (see Campos, Walle, Dahl, & Main, 2011; Gross & Thompson, 2007; Järvenoja & Järvenoja, 2009; Järvenoja, Volet, & Järvenoja; 2013; Mesquita, 2010; Mesquita & Boiger, 2014). Nevertheless, the theories and methods introduced provide useful scaffolding to move medical education research beyond an exclusive focus on stress and toward a comprehensive understanding of the rich array of emotions involved in learning and achievement.

References

Aldao, A., & Nolen-Hoeksema, S. (2010). Specificity of cognitive emotion regulation strategies: A transdiagnostic examination. *Behaviour Research and Therapy, 48,* 974–983.

Aldao, A., & Nolen-Hoeksema, S. (2012). When are adaptive strategies most predictive of psychopathology? *Journal of Abnormal Psychology, 121,* 276–281.

Arora, S., Sevdalis, N., Aggarwal, R., Sirimanna, P., Darzi, A., & Kneebone, R. (2010). Stress impairs psychomotor performance in novice laparoscopic surgeons. *Surgical Endoscopy and Other Interventional Techniques, 24,* 2588–2593.

Arora, S., Sevdalis, N., Nestel, D., Tierney, T., Woloshynowych, M., Kneebone, R. (2009). Managing intraoperative stress: What do surgeons want from a crisis training program? *The American Journal of Surgery, 197*, 537–543.

Artino, A. R. (2013). When I say…emotion in medical education. *Medical Education, 47,* 1062–1063.

Artino, A. R., Holmboe, E. S. & Durning, S. J. (2012). Control-value theory: Using achievement emotions to improve understanding of motivation, learning, and performance in medical education: *Medical Teacher 34*, 148–160.

Artino, A. R., La Rochelle, J. S., & Durning, S. J. (2010). Second year medical students' motivational beliefs, emotions, and achievement. *Medical Education*, *44*, 1203–1212.

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

Azenberg, Y., & Picard, R. (2014). Feel: A system for frequent event and electrodermal activity labeling. *IEEE Journal of Biomedical and Health Informatics, 18,* 266–277.

Azevedo, R., Harley, J., Trevors, G., Duffy, M., Feyzi-Behnagh, R., Bouchet, F., & Landis, R. S. (2013). Using trace data to examine the complex roles of cognitive, metacognitive, and emotional self-regulatory processes during learning with multi-agent systems. In R. Azevedo and V. Aleven (Eds.), *International handbook of metacognition and learning technologies* (pp. 427–449). NY: Springer.

Barrett, L. F. (2009). The future of psychology: Connecting mind to brain. *Perspectives in Psychological Science*, *4*, 326–339.

Becker, E. S., Goetz, T., Morger, V., & Ranellucci, R. (2014). The importance of teachers' emotions and instructional behavior for their students' emotions—An experience sampling analysis. *Teaching and Teacher Education, 43*, 15–26.

Ben-Eliyahu, A., & Linnenbrink, L. (2013). Extending self-regulated to include self-regulated emotion strategies. *Motivation and Emotion, 37*, 558–573.

Boekaerts, M. (2011). Emotions, emotion regulation, and self-regulation of learning. In B. Zimmerman & D. Schunk (Eds.). *Handbook of self-regulation of learning and performance* (pp. 408-425). New York: Routledge.

Bond, W. F., Deitrick, L. M., Arnold, D.C., Kostenbader, M., Barr, G. C., Kimmel, S. R., & Worrilow, C. C. (2004). Using simulation to instruct emergency medicine residents in cognitive forcing strategies. *Academic Medicine*, 79, 438–446.

Buss, D. M. (2014). Comment: Evolutionary criteria for considering an emotion "basic": Jealousy as an illustration. *Emotion Review, 4,* 313–331.

Calvo, R. A., & D'Mello, S. (2012). Frontiers of affect-aware learning technologies. IEEE *Intelligent Systems, 27*, 86–89.

Campos, J. J., Walle, E. A., Dahl, A., & Main, A. (2011). Reconceptualizing emotion regulation. *Emotion Review, 3*, 26–35.

Cheung, R. Y., & Au, T. K. (2011). Nursing students' anxiety and clinical performance. *Journal of Nursing Education, 50*, 286–289.

Chew, B. H., Zain, A. M., & Hassan, F. (2013). Emotional intelligence and academic performance in first and final year medical students: A cross-sectional study. *BMC Medical Education*, *13*, 1–10.

Clarke, S., Horeczko, T., Cotton, D., & Bair, A. (2014). Heart rate, anxiety and performance of residents during a simulated critical clinical encounter: A pilot study. *BMC Medical Education, 14*, 1–8.

Craig, S., D., D'Mello, S., Witherspoon A., & Graesser, A. (2008). Emote-aloud during learning with AutoTutor: Applying the Facial Action Coding System to cognitive-affective states during learning. *Cognition and Emotion, 22*, 777–788.

Croskerry, P. (2003). Cognitive forcing strategies in clinical decision-making. *Annals of Emergency Medicine, 41*, 110–120.

Croskerry, P., Abbass, A. A., & Wu, A. W. (2008). How doctors feel: Affective issues in patients' safety. *The Lancet, 372,* 1205–1206.

Csikszentmihalyi, M., & Larson, R. (1987). Validity and reliability of the experience sampling method. *Journal of Nervous and Mental Diseases, 9*, 526–536.

Darwin, C. (1872). *The expression of the emotions in man and animals*. London, UK: John Murray.

de Vries, E. N., Ramrattan, M. A., Smorenburg, S. M., Gouma, D. J., & Boermeester, M. A. (2008). The incidence and nature of in-hospital adverse events: A systematic review. *Quality and Safety in Health Care, 3,* 216–223.

D'Mello, S. (2013). A selective meta-analysis on the relative incidence of discrete affective states during learning with technology. *Journal of Educational Psychology, 105,* 1082–1099.

D'Mello, S. K., Dale, R. A., & Graesser, A. C. (2012). Disequilibrium in the mind, disharmony in the body. *Cognition and Emotion, 26*, 362–374.

D'Mello, S. K., & Graesser, A. C. (2009). Automatic detection of learners' emotions from gross body language. *Applied Artificial Intelligence, 23*(2), 123-150.

D'Mello, S. K., & Graesser, A. C. (2010). Multimodal semi-automated affect detection from conversational cues, gross body language, and facial features. *User Modeling and User-adapted Interaction, 20*(2), 147–187.

D'Mello, S. K., & Graesser, A. C. (2012). Dynamics of affective states during complex learning. *Learning and Instruction, 22,* 145–157.

D'Mello, S., Lehman, B., Pekrun, R., & Graesser, A. (2014). Confusion can be beneficial for learning. *Learning and Instruction, 29,* 153–170.

Duffy, M. C., Azevedo, R., Sun, N., Griscom, S., Stead, V., Crelinsten, L., ... Lachapelle, K. (2015a). Team regulation in a simulated medical emergency: An in-depth analysis of cognitive, metacognitive, and affective processes. *Instructional Science, 43*, 401–426.

Duffy, M. C., Lajoie, S., Jarrell, A., Pekrun, R., Azevedo, R., & Lachapelle, K. (2015b, April). *Emotions in medical education: Developing and testing a scale of emotions across medical learning environments*. Paper presented at the annual meeting of the American Educational Research Association, Chicago, IL.

Duffy, M. C., Lajoie, S., Pekrun, R., Ibrahim, M., & Lachapelle, K. (2015c, March). *Simulation versus operating room: Examining relations between emotional states, learning processes, and technical skills performance across training environments*. Poster presented at the annual meeting of the consortium of American College of Surgeons Accredited Education Institutes, Chicago, IL.

Dunn, B. D., Evans, D., Makarova, D., White, J., & Clark, L. (2012). Gut feelings and the reaction to perceived inequity: The interplay between bodily responses, regulation, and perception shapes the rejection of unfair offers on the ultimatum game. *Cognitive, Affective, & Behavioral Neuroscience, 12*, 149–429.

Dyrbye, L. N., West, C. P., Satele, D., Boone, S., Tan, L., Sloan, J., & Shanfelt, T. D. (2014). Burnout among U.S. medical students, residents, and early career physicians relative to the general U.S. population. *Academic Medicine, 89,* 443–451.

Efklides, A. (2011). Interactions of metacognition with motivation and affect in self-regulated learning: The MASRL model. *Educational Psychologist, 46*(1), 6–25.

Eisenberg, N., Fabes, R. A., Guthrie, I. K., & Reiser, M. (2000). Dispositional emotionality and regulation: Their role in predicting quality of social functioning. *Journal of Personality and Social Psychology, 78*, 136–157.

Ekman, P. (1972). Universal and cultural differences in facial expressions of emotions. In J. K. Cole (Ed.), *Nebraska symposium on motivation, 1971* (pp. 207–283). Lincoln, NE: University of Nebraska Press.

Ekman, P. (1992). An argument for basic emotions. *Cognition and Emotion*, *6*, 169-200.

Ekman, P. (1999). Basic emotions. In T. Dalgleish, & M. Power (Eds.), *Handbook of Cognition and Emotion* (pp. 45–60). Sussex, UK: John Wiley & Sons.

Ekman, P., & Cordar, D. (2011). What is meant by calling emotion basic. *Emotion Review, 3,* 364–370.

Ekman, P., Friesen, W. V., & Hager, J. C. (2002). *Facial action coding system* [E-book]. Salt Lake City, UT: Research Nexus.

Ellsworth, P. C. (2013). Appraisal theory: Old and new questions. *Emotion Review*, *5*, 125–131.

Emotion [Def. 2c]. (n.d.). *Merriam-Webster Online*. Retrieved September 24, 2014, from http://www.merriam-webster.com/dictionary/emotion.

Ericsson, K., & Simon, H. (1993). *Protocol analysis: Verbal reports as data* (2nd ed.). Boston: MIT Press.

Fedor, S., & Picard, R. (2014). Ambulatory EDA: *Comparisons of bilateral forearm and calf locations*. Poster presented at the 54th Annual Meeting of the Society of Psychophysiological Research. Atlanta, USA.

Fisher, A. H., & van Kleef, G. A. (2010). Where have all the people gone? A plea for including social interaction in emotion research. *Emotion Review, 2,* 208–211.

Fontaine, J. R. J., Scherer, K. R., Soriano, C. (2013). *Components of emotional meaning: A sourcebook*. Oxford: Oxford University Press.

Fraser, K., Huffman, J., Ma, I., Sobczak, M., McIlwrick, J., Wright, B., & McLaughlin, K. (2014). The emotional and cognitive impact of unexpected simulated patient death: A randomized controlled trial. *Chest, 145*, 958–963.

Fraser, K., Ma, I., Teteris, E., Baxter, H., Wright, B., & McLaughlin, K. (2012). Emotion, cognitive load and learning outcomes during simulation training. *Medical Education, 46*, 1055–1062.

Frijda, N. H. (1986). *The emotions*. New York, NY: Cambridge University Press.

Frijda, N. H. (1987). Emotion, cognitive structure, and action tendency. *Cognition and Emotion*, *1*, 115–143.

Frijda, N. H. 2007. *The laws of emotion*. Mahwah, N.J.: L. Erlbaum Associates.

Goetz, T., Frenzel, A., Hall, N. C., Nett, U. E., Pekrun, R., & Lipnevich, A. A. (2014). Types of boredom: An experience sampling approach. *Motivation & Emotion, 38,* 401–419.

Goetz, T., Frenzel, A. C., Pekrun, R., Hall, N. C., & Lüdtke, O. (2007). Between- and within-domain relations of students' academic emotions. *Journal of Educational Psychology, 99*, 715–733.

Goetz, T., Pekrun, R., & Hall, N., & Haag, L. (2006). Academic emotions from a social-cognitive perspective: Antecedents and domain specificity of students' affect in the context of latin instruction. *The British Psychological Society, 76*, 289–308.

Graesser, A., D'Mello, S., & Strain, A. (2014). Emotions in advanced learning technologies. In R. Pekrun & L. Linnenbrink-Garcia (Eds.), *International handbook of emotions in education* (pp. 289–310). New York: Routledge.

Gross, J. J. (1998a). Antecedent- and response-focused emotion regulation: Divergent consequences for experience, expression, and physiology. *Journal of Personality and Social Psychology, 74,* 224–237.

Gross, J. J. (1998b). The emerging field of emotion regulation: An integrative review. *Review of General Psychology, 2,* 271–299.

Gross, J. J. (2001). Emotion regulation in adulthood: Timing is everything. *Current Directions in Psychological Science, 10,* 214–219.

Gross, J. J., & Barrett, L. F. (2011) Generation and emotion regulation: One or two depends on your point of view. *Emotion Review, 3,* 8–16.

Gross, J. J., & John, O. P. (2003). Individual differences in two emotion regulation processes: Implications for affect, relationships and well-being. *Journal of Personality and Social Psychology, 85,* 348–362.

Gross, J. J., Sheppes, G., & Urry, H. L. (2011). Emotion generation and emotion: A distinction we should make (carefully), *Cognition and Emotion, 25*, 765–781.

Gross, J. J., & Thompson, R. A. (2007). Emotion regulation: Conceptual foundations. In J. J. Gross (Ed.), *Handbook of emotion regulation* (pp. 3–24). New York, NY: Guilford Press.

Harvey, A., Bandiera, G., Nathens, A. B., & LeBlanc, V. R. (2012). Impact of stress on resident performance in simulated trauma scenarios. *Journal of Trauma Acute Care Surgery 72,* 497–503.

Hassan, I., Weyers, P., Maschuw, K., Dick, B., Gerdes, B., Rothmund, M., & Zielke, A. (2006). Negative stress-coping strategies among novices in surgery correlate with poor virtual laparoscopic performance. *British Journal of Surgery, 93*, 1554–1559.

Hektner, J. M., Schmidt, J. A., & Csikszentmihalyi, M. (2007). Experience sampling method: Measuring the quality of everyday life. Thousand Oaks, CA: Sage Publications, Inc.

Hunziker, S., Laschinger, L., Portmann-Schwarz, S., Semmer, N.K., Tschan, F., & Marsch, S. (2011). *Perceived stress and team performance during a simulated resuscitation. Intensive Care Medicine, 37,* 1473–1479.

Hunziker, S., Pagani, S., Fasleer, K., Tschan, F., Semmer, N. K., & Marsch, S. (2013). Impact of a stress coping strategy on perceived stress levels and performance during a simulated cardiopulmonary resuscitation. *BMC Emergency Medicine, 13,* 8–17.

Hunziker, S., Semmer, N.K., Tschan, F., Schuetz, P., Mueller, B., & Marsch, S. (2012).

Dynamics and association of different acute stress markers with performance during a simulated resuscitation. *Resuscitation, 83*, 572–578.

Immordino-Yang, M. H., & Christodoulou, J. A. (2014). Neuroscientific contributions to understanding and measuring emotions in educational contexts. In R. Pekrun & L. Linnenbrink-Garcia (Eds.), *International handbook of emotions in education* (pp. 607–624). New York: Routledge

Izard, C. E. (1993). Four systems for emotion activation: Cognitive and noncognitive processes. *Psychological Review*, *100*, 68–90.

Izard, C. E. (2007). Basic emotions, natural kinds, emotion schemas, and a new paradigm. *Perspectives on Psychological Science*, *2*, 260–280.

Izard, C. E. (2010). The many meanings/aspects of emotion: Definitions, functions, activation, and regulation, *Emotion Review, 2,* 363–370.

Izard, C. E. (2011). Forms and functions of emotions: Matters of emotion–cognition interactions. *Emotion Review*, *3*, 371–378.

James, W. (1890). *The principles of psychology*. New York, NY: Dover.

Järvenoja, H., Volet, S., Järvelä, S. (2013). Regulation of emotions in socially challenging learning situations: An instrument to measure the adaptive and social nature of the regulation process. *International Journal of Experimental Educational Psychology, 33*, 31–58.

Kasman, D. L., Fryer-Edwards, K., & Braddock, C. H. (2003). Educating for professionalism: Trainees' emotional experiences on IM and pediatrics inpatient ward. *Academic Medicine, 78,* 730–741.

Keitel, A., Ringleb, M., Schwartges, I., Weik, U., Picker, O., Stockhorst, U., & Deinzer, R. (2011). Endocrine and psychological stress responses in a simulated emergency situation. *Psychoneuroendocrinology, 36*, 98–108.

Keltner, D., & Buswell, B. N. (1997). Embarrassment: Its distinct form and appeasement functions. *Psychological Bulletin, 122,* 250–270.

Kohn, K. T., Corrigan, J. M., & Donaldson, M. S. (2000). *To err is human: Building a safer health system*. Washington, DC: National Academy Press.

Kreibig, S. D., & Gendolla, G. H. E. (2014). Autonomic nervous system measurement of emotion in education and achievement settings. In R. Pekrun, & L. Linnenbrink-Garcia

(Eds.), *International handbook of emotions in education* (pp. 625–642). New York, NY: Routledge.

Kring, A. M., & Sloan, D. M. (2007). The facial expression coding system (FACES): Development, validation, and utility. *Psychological Assessment, 19,* 210–224.

Lajoie, S. P. (2009). Developing professional expertise with a cognitive apprenticeship model: Examples from Avionics and Medicine. In K. A. Ericsson (Ed.). *Development of professional expertise: Toward measurement of expert performance and design of optimal learning environments* (pp. 61–83). New York: NY. Cambridge University Press.

Lajoie, S., Poitras, E., Naismith, L., Gauthier, G., Summerside, C., & Kazemitabar, M. (2013). Modelling domain-specific self-regulatory activities in clinical reasoning. In H. C. Lane, K. Yacef, & J. Mostow (Eds.), *International artificial intelligence and education proceedings* (pp. 632–635). Amsterdam: IOS Press.

Larsen, J. T., & McGraw, P. (2011). Further evidence for mixed emotions. *Journal of Personality and Social Psychology, 100*, 1095–1110.

Lazarus, R. S. (1966). *Psychological stress and the coping process*. New York, NY: McGraw-Hill.

Lazarus, R. S. (1991). *Emotion and adaptation*. New York, NY: Oxford University Press.

Lazarus R. S., & Folkman, S. (1984). *Stress, appraisal, and coping*. New York: Springer.

Leonard, M., Graham, S. & Bonacum, D. (2004). The human factor: The critical importance of effective teamwork and communication in providing safe care. *Quality and Safety in Health Care,* 13, 85–90.

Lingard, L., Regehr, G., Orser, B., Reznick, R., Baker, G. R. Doran, D., Espin, S., Bohnen, J., & Whyte, S. (2008). Evaluation of a preoperative checklist and team briefing among surgeons, nurses, and anesthesiologists to reduce failures in communication. *Archives of Surgery 143*, 12–17.

LeBlanc, V. R. (2009). The effects of acute stress on performance: implications for health professions education. *Academic Medicine, 84,* 25–33.

LeBlanc, V. R., & Bandier, G. W. (2007). The effects of examination stress on the performance of emergency medicine residents. *Medical Education, 41,* 556-564.

LeBlanc, V. R., MacDonald, R. D., McArthur, B., King, K., & Lepine, T. (2005). Paramedic performance in calculating drug dosages following stressful scenarios in a human patient simulator. *Prehospital Emergency Care, 9,* 439–444.

LeBlanc, V. R., Manser, T., Weinger, M. B., Musson, D., Kutzin, J., & Howard, S. K. (2011). The study of factors affecting human and systems performance in healthcare using simulation. *Simulation in Healthcare, 6,* 24–29.

LeBlanc, V. R., McConnell, M. M., & Monteiro, S. D. (2014). Predictable chaos: a review of the effects of emotions on attention, memory and decision-making. *Advances in Health Sciences Education, 20*, 265–282.

LeBlanc, V. R., Regehr, C., Tavares, W., Scott, A. K., Macdonald, R., & King, K. (2012). The impact of stress on paramedic performance during simulated critical events. *Prehospital and Disaster Medicine, 27*, 369–374.

LeBlanc, V. R., Woodrow, S. I., Sidhu, R., & Dubrowski, A. (2008). Examination stress leads to improvements on fundamental technical skills for surgery. *American Journal of Surgery, 196,* 114–119.

LeDoux, J. E. (2014). Comment: What's basic about the brain mechanisms of emotions? *Emotion Review 6*, 318–320.

Manser, T., & Foster, S. (2011). Effective handover communication: An overview of research and improvement efforts. *Safety in Anaesthesia 25,* 181–91.

Matejka, M., Kazzer, P., Seehausen, M., Bajbouj, M., Klann-Delius, G., Menninghaus, W., Jacobs, A. M., Heekeren, H. R., & Prehn, K. (2013). Talking about emotion: Prosody and skin conductance indicate emotion regulation. *Frontiers in Psychology, 4,* 1–11.

Matsumoto, D., & Hwang, H. S. (2011). Evidence for training the ability to read microexpressions of emotion. *Motivation and Emotion*, *35*, 181–191.

Mauss, I. B., Cook, C. K., & Gross, J. J. (2007). Automatic emotion regulation during anger provocation. *Journal of Experimental Social Psychology, 43,* 698–711.

Mauss, I. B., & Robinson, M. D. (2009). Measures of emotions: A review. *Cognition and Emotion, 23*, 209–237.

McConnell, M. M., & Eva, K. W. (2012). The role of emotion in the learning and transfer of clinical skills and knowledge. *Academic Medicine, 87,* 1316–1322.

McWilliams, L. A., Cox, B. J., & Enns, M. W. (2003). Use of coping inventory for stressful situations in a clinically depressed sample: Factor structure, personality correlates, and prediction of distress. *Journal of Clinical Psychology, 59*, 423–437.

Meinhardt, J., & Pekrun, R. (2003). Attentional resource allocation to emotional events: An ERP study. *Cognition and Emotion, 17*, 477–500.

Mesquita, B. (2010). Emoting: A contextualized process. In B. Mesquita, L. F. Barrett & E. Smith (Eds.), *The mind in context* (pp. 83–104). New York, NY: Guilford.

Mesquita, B., & Boiger, M. (2014). Emotions in context: A sociodynamic model of emotions. *Emotion Review*, *6*, 298–302.

Meunier, J., Merckaert, I., Libert, Y., Delvaux, N., Etienne, A.-M., Lienard, A., Bragard, I., Marchal, S., Reynaert, C., Slachmuylder, J.-L., & Razavi, D. (2013). The effect of communication skills training on residents' physiological arousal in a breaking bad news simulated task. *Patient Education & Counseling, 93,* 40-47.

Mitmasgruber, H., Beck, T. N., Höfer, S., & Schübler, G. (2009). When you don't like what you feel: Experiential avoidance, mindfulness, and meta-emotion in emotion regulation. *Personality and Individual Differences, 46,* 448–453.

Moors, A. (2014). Flavors of appraisal theories of emotion. *Emotion Review, 6*, 303-307.

Moors, A., Ellsworth, P. C., Scherer, K. R., & Frijda, N. H. (2013). Appraisal theories of emotion: State of the art and future development. *Emotion Review, 5*, 119–124.

Muis, K. R., Duffy, M. C., Trevors, G., & Ranellucci, J., & Foy, M. (2014). What were they thinking? Using cognitive interviewing to examine the validity of self-reported epistemic beliefs. *International Journal of Education, 2,* 17–32.

Müller, M. P., Hansel, M., Fichtner, A., Hardt, F., Weber, S., Kirschbaum, C., Rüder, S., Walcher, F., Koch, T., & Eich, C. (2009). Excellence in performance and stress reduction during two different full scale simulator training courses: A pilot study. *Resuscitation, 80,* 919–924.

Mulligan, K., & Scherer, K. R. (2012). Toward a working definition of emotion. *Emotion Review, 4,* 345–357.

Murphy, F. C., Nimmo-Smith, I., & Lawrence, A. D. (2003). Functional neuroanatomy of emotions: a meta-analysis. *Cognitive Affective & Behavioral Neuroscience, 3,* 207–33.

Nesse, R. M. (2014). Comment: A general "theory of emotion" is neither necessary nor possible. *Emotion Review*, *6*, 320–322.

Nesse, R. M., & Ellsworth, P. C. (2009). Evolution, emotions, and emotional disorders. *American Psychologist*, *6*, 129–139.

Nett, U. E., Goetz, T., & Hall, N. C. (2011). Coping with boredom in school: An experience sampling perspective. *Contemporary Educational Psychology, 36*, 49–59.

O'Callaghan, A. (2013). Emotional congruence in learning and health encounters in medicine: Addressing as aspect of the hidden curriculum. *Advances in Health Sciences Education*, *18,* 305–317.

Öhman, A., Hamm, A., & Hugdahl, K. (2000). Cognition and the autonomic nervous system: Orienting, anticipation, and conditioning. In J. T. Cacioppo, L. G. Tassinary, & G. G. (Eds.), *Handbook of psychophysiology* (Vol. 2, pp. 533–575). New York: Cambridge University Press.

Ortony, A., & Turner, T. J. (1990). What's basic about basic emotions? *Psychological Review*, *97*, 315–331.

Panksepp, J., & Watt, D. What is basic about basic emotions? Lasting lessons from affective neuroscience. *Emotion Review, 3,* 387–396.

Pantic, M., & Rothkrantz, L. J. M. (2000). Automatic analysis of facial expressions: The state of the art. *Pattern Analysis and Machine Intelligence, IEEE Transactions on, 22*(12), 1424–1445.

Pekrun, R. (1992). The impact of emotions on learning and achievement: Towards a theory of cognitive/motivational mediators. *Applied Psychology, 41*, 359–376.

Pekrun, R. (2006). The control-value theory of achievement emotions: Assumptions, corollaries, and implications for educational research and practice. *Educational Psychology Review, 18*, 315–341.

Pekrun, R., & Bühner, M. (2014). Self-report measures of academic emotion. In R. Pekrun & L. Linnenbrink-Garcia (Eds.), *International handbook of emotions in education* (pp. 561–579). New York: Routledge

Pekrun, R., Elliot, A. J., & Maier, M. A. (2009). Achievement goals and achievement emotions: Testing a model of their join relations with academic performance. *Journal of Educational Psychology, 101,* 115-135.

Pekrun, R., Frenzel, A. C., Goetz, T., & Perry, R. P. (2007). The control-value theory of achievement emotions: An integrative approach to emotions in education. In P. A. Schutz & R. Pekrun (Eds.), *Emotion in education* (pp. 13–36). San Diego, CA: Academic Press.

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,* 36–48.

Pekrun, R., Goetz, T., Titz, W., & Perry, R. P. (2002). Academic emotions in students' self-regulated learning and achievement: A program of qualitative and quantitative research. *Educational Psychologist*, *37*, 91–105.

Pekrun, R., & Linnenbrink-Garcia, L. (2014). Introduction to emotions in education. In R. Pekrun & L. Linnenbrink-Garcia (Eds.), *International handbook of emotions in education* (pp. 1–10). New York: Francis & Taylor/Routledge.

Pekrun, R., Meier, E., Muis, K. R., & Sinatra, G. M. (revise and resubmit). Measuring emotions related to knowledge and knowing: The Epistemic Emotions Scales (ESS). *Cognition and Emotion*.

Pekrun, R., & Perry, R. P. (2014). Control-value theory of achievement emotions. In R. Pekrun & L. Linnenbrink-Garcia (Eds.), *International handbook of emotions in education* (pp. 120–141). New York: Francis & Taylor/Routledge.

Pekrun, R., & Stephens, E. J. (2012). Academic emotions. In K. R. Harris, S. Graham, T. Urdan, S. Graham, J. M. Royer, & M. Zeidner (Eds.), *APA educational psychology handbook* (Vol. 2, pp. 3–31). Washington, DC: American Psychological Association.

Picard, R. W. (2010). Emotion research by the people, for the people. *Emotion Review, 2*, 250–254.

Pintrich, P. R. (2000). The role of goal orientation in self-regulated learning. In M. Boekaerts, P. Pintrich, & M. Zeidner (Eds.), *Handbook of self-regulation* (pp. 451–502). San Diego, CA: Academic Press.

Pintrich, P. (2004). A conceptual framework for assessing motivation and self-regulated learning in college students. *Educational Psychology Review, 16,* 385–407.

Piquette, D., Reeves, S., & LeBlanc, V. R. (2009). Stressful intensive care unit medical crises: How individual responses impact on team performance. *Critical Care Medicine, 37*, 1251–1255.

Piquette, D., Tarshis, J., Sinuff, T., Pinto, R., & LeBlanc, V. R. (2014). Impact of acute stress on resident performance during simulated resuscitation episodes: a perspective randomized cross-over study. *Teaching and Learning in Medicine: An International Journal, 26,* 9–16.

Plutchik, R. (1980). A general psychoevolutionary theory of emotion. In R. Plutchik & H. Kellerman (Eds.), *Emotion: Theory, research, and experience: Vol. 1. Theories of emotion* (pp. 3–33). New York: Academic.

Plutchik, R. (2001). The nature of emotions. *American Scientist*, *89*, 344–350.

Poh, M., Swenson, N. C., & Picard, R., W. (2010). A wearable sensor for unobtrusive, long-term assessment of electrodermal activity. *IEEE*, *Transactions on Biomedical Engineering*, 57, 1243–1252.

Reisenzein, R., Junge, M., Studtmann, M., & Huber, O. (2014). Observational approaches to the measurement of emotions. In R. Pekrun & L. Linnenbrink-Garcia (Eds.), *International handbook of emotions in education* (pp. 580–606). New York: Routledge.

Russell, J. A. (2003). Core affect and the psychological construction of emotion. *Psychological Review*, *110*, 145–172.

Russell, J. A., & Barrett, L. F. (1999). Core affect, prototypical emotional episodes, and other things called emotion: Dissecting the elephant. *Journal of Personality and Social Psychology*, *7*, 805–819.

Satterfield, J. M., & Hughes, E. (2007). Emotion skills training for medical students: A systematic review. *Medical Education, 41,* 935–941.

Scarantino, A., & Griffiths, P. (2011). Don't give up on basic emotions. *Emotion Review, 3,* 444–454.

Schachter, S., & Singer, J. E. (1962). Cognitive, social and physiological determinants of emotional state. *Psychological Review*, *69*, 379–399.

Scherer, K. R. (1984). Emotion as a multicomponent process: A model and some cross-cultural data. In P. Shaver (Ed.), *Review of personality and social psychology* (Vol. 5, pp. 37–63). Beverly Hills, CA: Sage.

Scherer, K. (2005). What are emotions? And how can they be measured? *Social Science Information, 44*, 695–729.

Scherer, K. R. (2009). The dynamic architecture of emotion: Evidence for the component process model. *Cognition and Emotion, 23,* 1307–1351.

Scherer, K. R. (2013). The nature and dynamics of relevance and valence appraisals: Theoretical advances and recent evidence. *Emotion Review*, *5*, 150–162.

Scherer, K. R., Shuman, V., Fontaine, J. R. J., & Soriano, C. (2013). The GRID meets the wheel: Assessing emotional feeling via self-report. In J. R. J. Fontaine, K. R. Scherer, & C. Soriano (Eds.), *Components of emotional meaning: A sourcebook* (pp. 281–298). Oxford, United Kingdom: Oxford University Press.

Shapiro, J. (2011). Does medical education promote professional alexithymia? A call for attending to the emotions of patients and self in medical training. *Academic Medicine, 86,* 326–332.

Sharma, N., & Gedeon, T. (2012). Objective measures, sensors and computational techniques for stress recognition and classification: A survey. *Computer Methods and Programs in Biomedicine, 103,* 1287–1301.

Shuman, V., & Scherer, K. R. (2014). Concepts and structure of emotions. In R. Pekrun & L. Linnenbrink-Garcia (Eds.), *International handbook of emotions in education* (pp. 13–35). New York: Routledge.

Smith, C., A., & Ellsworth, P. C. (1985). Patterns of cognitive appraisal in emotion. *Journal of Personality and Social Psychology, 4*, 813–838.

Solomon, R. C. (2003). *Not passion's slave: Emotions and choice*. New York, NY: Oxford University Press.

Spielberger, C. D., Gorsuch, R., Lushene, R., Vagg, P., & Jacobs, G. (1983). *Manual for the State-Trait Anxiety Inventory*. Palo Alto, CA: Consulting Psychologists Press.

Tamir, M. (2011). The maturing field of emotion regulation. *Emotion Review, 3*, 3–7.

Tracy, J. (2014a). An evolutionary approach to understanding distinct emotions. *Emotion Review*, *6*, 308–312.

Tracy, J. (2014b). Author reply: Incompatible conclusions or different levels of analysis? *Emotion Review*, *6*, 330–331.

Tracy, J. L., & Randles, D. (2011). Four models of basic emotions: A review of Ekman and Cordaro, Izard, Levenson, and Panksepp and Watt. *Emotion Review, 3,* 397–405.

Tracy, J. L., & Robins, R. W. (2004). Show your pride: Evidence for a discrete emotion expression. *Psychological Science, 15,* 194–197.

Tracy, J. L., & Robins, R. W. (2007). The prototypical pride expression: Development of a nonverbal coding scheme. *Emotion, 7*, 789–801.

Tugade, M. M., & Fredrickson, B. L. (2007). Regulation of positive emotions: Emotion regulation strategies that promote resilience. *Journal of Happiness Studies, 8,* 311–333.

Turner, J. C., & Trucano, M. (2014). Measuring situated emotion. In R. Pekrun & L. Linnenbrink-Garcia (Eds.), *International handbook of emotions in education (*pp. 643–658). New York: Routledge.

van Dulmen, S., Tromp, F., Grosfeld, F., ten Cate, O., & Bensing, J. (2007). The impact of assessing simulated bad news consultations on medical students' stress response and communication performance. *Psychoneuroendocrinology, 32*, 943–95.

Van Gog, T., & Jarodzka, H. (2013). Eye tracking as a tool to study and enhance cognitive and metacognitive processes in computer-based learning environments. In R. Azevedo & V. Aleven (Eds.), *International handbook of metacognition and learning technologies* (pp. 143–156). New York, NY: Springer.

Webb, T. L., Miles, E., & Sheeran, P. (2012). Dealing with feeling: A meta-analysis of the effectiveness of strategies derived from the process model of emotion regulation. *Psychological Bulletin, 138,* 775–808.

Weiner, B. (2007). Examining emotional diversity on the classroom: An attribution theorist considers the moral emotions. In P. A. Shutz & R. Pekrun (Eds.), *Emotion in education* (pp. 73–88). San Diego, CA: Academic Press.

Wetzel, C. M., Black, S. A., Hanna, G. B., Athanasiou, T., Kneebone, R. L., Nestel, D., Wolfe, J. H. N., & Woloshynowych, M. (2010). The effects of stress and coping on surgical performance during simulations. *Annals of Surgery, 251*, 171–176.

Wetzel, C. M., Kneebone, R. L., Woloshynowych. M., Nestel, D., Moorthy, K., Kidd, J., & Darzi, A. (2006). The effects of stress on surgical performance, *American Journal of Surgery*, *191,* 5–10.

Winne, P. H. (2001). Self-regulated learning viewed from models of information processing. In B. J., Zimmerman & D. H. Schunk (Eds.), *Self-regulated learning and academic achievement: Theoretical perspectives* (2nd Ed) (pp. 153–189). Mahwah, NJ: Erlbaum.

Winne, P., & Hadwin, A. (1998). Studying as self-regulated learning. In D. Hacker, J. Dunlosky, & A. Graesser (Eds.), *Metacognition in educational theory and practice* (pp. 227–304). Mahwah, NJ: Erlbaum.

Winne, P., & Hadwin, A. (2008). The weave of motivation and self-regulated learning. In D. Schunk & B. Zimmerman (Eds.), *Motivation and self-regulated learning: Theory, research, and applications* (pp. 297–314). Mahwah, NJ: Erlbaum.

Yik, M., Russell, J. A., & Steiger, J. H. (2011). A 12-point circumplex structure of core affect. *Emotion, 11,* 705–731.

Zeng, Z., Pantic, M., Roisman, G. I., & Huang, T.S. (2009). A survey of affect recognition methods: Audio, visual, and spontaneous expressions. *IEEE Transactions on Pattern Analysis and Machine Learning, 31*, 39–58.

Zentner, M., Grandjean, D., & Scherer, K. (2008). Emotions evoked by the sound of music: Characterization, classification, and measurement. *Emotion*, *8*, 494–521.

Zimmerman, B. (2000). Attaining self-regulation: A social cognitive perspective. In M. Boekaerts, P. Pintrich, & M. Zeidner (Eds.), *Handbook of self-regulation* (pp. 13–39). San Diego, CA: Academic Press.

Zimmerman, B. (2011). Motivational sources and outcomes of self-regualted learning and performance. In B. J. Zimmerman & D. H. Schunk (Eds.), *Handbook of self-regulation of learning and performance* (pp. 49–64). New York: Routledge.

Zeidner, M. (2014). Anxiety in education. In R. Pekrun & L. Linnenbrink-Garcia (Eds.), *International handbook of emotions in education* (pp. 265–288). New York: Routledge.

Table 1

Examples of empirical studies examining emotions within medical and health sciences education

Authors	Title	Emotions	Theory	Measure(s)
Arora et al (2010)	Stress impairs psychomotor performance in novice laparoscopic surgeon	Stress/anxiety	Cognitive appraisal theory of stress	Self-report (questionnaire) Physiological (cortisol level, heart rate)
Artino et al (2010)	Second-year medical students' motivational beliefs, emotions, and achievement	Positive and negative emotions	Control-value theory	Self-report (questionnaire)
Cheung & Au (2011)	Nursing students' anxiety and clinical performance	Stress/anxiety	Not explicitly stated	Self-report (questionnaire)
Clarke et al (2014)	Heart rate, anxiety, and performance during a simulated critical clinical encounter: a pilot study	Stress/anxiety	Cognitive appraisal theory of stress	Self-report (questionnaire) Physiological (heart rate)
Duffy et al (2015a)	Team regulation in a simulated medical emergency: An in-depth of analysis of cognitive, metacognitive, and affective processes	Positive and negative emotions	Control-value theory	Self-report (interview) Behavioral (body language, speech, facial expression)
van Dulmen et al (2007)	The impact of assessing simulated bad news on medical students' stress response and communication performance	Stress/anxiety	Not explicitly stated	Self-report (questionnaire) Physiological (cortisol level, heart rate, blood pressure)
Fraser et al (2012)	Emotion, cognitive load, learning outcomes during simulation training	Positive and negative emotions	Circumplex model	Self-report (questionnaire)
Fraser et al (2014)	The emotional and cognitive impact of unexpected simulated patient death	Positive and negative emotions	Circumplex model	Self-report (questionnaire)
Harvey et al (2012)	Impact of stress of resident performance in simulated trauma scenarios	Stress/anxiety	Cognitive appraisal theory of stress	Self-report (questionnaire) Physiological (cortisol level, heart rate)
Hunziker et al (2011)	Perceived stress and team performance during a simulated resuscitation	Positive and negative emotions	Not explicitly stated	Self-report (questionnaire)
Hunziker et al (2012)	Dynamics and association of different acute stress markers with performance during a simulated resuscitation	Stress/anxiety	Cognitive appraisal theory of stress	Self-report (questionnaire) Physiological (cortisol level, heart rate)
Keitel et al (2011)	Endocrine and psychological stress responses in a simulated emergency situation	Stress/anxiety	Not explicitly stated	Self-report (questionnaire) Physiological (cortisol level)
LeBlanc et al (2005)	Paramedic performance in calculating drug dosages following stressful scenarios in a human patient simulator	Stress/anxiety	Not explicitly stated	Self-report (questionnaire)
LeBlanc et al (2008)	Examination stress leads to improvements on fundamental technical skills for surgery	Stress/anxiety	Cognitive appraisal theory of stress	Self-report (questionnaire)

LeBlanc et al (2012)	The Impact of Stress on Paramedic Performance During Simulated Critical Events	Stress/anxiety	Cognitive appraisal theory of stress	Self-report (questionnaire) Physiological (cortisol level
LeBlanc & Bandiera (2007)	The effects of examination stress on the performance of emergency medicine residents	Stress/anxiety	Not explicitly stated	Self-report (questionnaire)
Meunier et al (2013)	The effect of communication skills training on residents' physiological arousal in a breaking bad news simulated task	Stress/anxiety	Cognitive appraisal theory of stress	Self-report (questionnaire) Physiological (cortisol level, heart rate)
Müller et al (2009)	Excellence in performance and stress reduction during two different full scale simulator training courses: A pilot study	Stress/anxiety	Cognitive appraisal theory of stress	Physiological (cortisol level)
Piquette et al (2009)	Stressful intensive care unit medical crises: How individual responses impact on team performance	Stress/anxiety	Cognitive appraisal theory of stress	Self-report (interview)
Piquette et al (2014)	Impact of acute stress on resident performance during simulated resuscitation episodes: A prospective randomized cross-over study	Stress/anxiety	Cognitive appraisal theory of stress	Self-report (questionnaire) Physiological (cortisol level)
Wetzel et al (2006)	The effects of stress on surgical performance	Stress/anxiety	Not explicitly stated	Self-report (interview)
Wetzel et al (2010)	The effects of stress and coping on surgical performance during simulations	Stress/anxiety	Cognitive appraisal theory of stress	Self-report (questionnaire) Behavioral (not specified) Physiological (cortisol levels, heart rate)

Note. Studies were included in this review if they: (1) measured state emotions/short-term affective states (e.g., acute stress) rather than trait emotions/long-term affective states (e.g., burnout) and (2) examined emotions in relation to learning and/or performance variables.

Table 2

Summary of review of emotion theories and applications for medical education

Theory	Core tenants	Advantages	Disadvantages/gaps	Medical education questions
Basic emotions theories	• Primitive forms of emotions are universal • Evolutionary purpose and biological basis of emotions • Emotions are innate and automated • Discrete categories	• Introduces nature and structure of emotions • Adaptive function of emotions • Evidence mapping emotion to facial expression • Evidence linking emotions to brain regions	• Lacking clear criteria and taxonomy for basic emotions • Overlooks role of socialization process • Overlooks role of appraisals in emotion generation • Does not effectively incorporate range of emotional experience	• Are certain emotions universal across medical education (med ed) environments? • Are basic emotions more intense among novices than experts? • Are basic emotions more highly activated in crisis/high-stakes environments/tasks?
Appraisal theories of emotions	• Emotions are multi-componential • Emotions are continuous processes or episodes • Appraisals differentiate emotions • Appraisals range from conscious to unconscious	• Accounts for variability in emotional experience across individuals and events • Supports a rich array of nuanced emotions that are likely to occur in adulthood	• Unclear how many and which appraisals are needed to predict variability in emotions • Mechanisms underlying appraisal process unclear • Degree of automation involved in appraisal process unclear	• How are emotions experienced through multiple components among medical trainees? • Is there alignment among components? • Which appraisal combinations predict emotions in med ed? • How might emotions change over time, task, and medical trainee?
Control-value theory of emotions	• Control and value appraisals considered most important for predicting emotions • Classifies achievement emotions according to valence, arousal, object-focus • Differentiates between activity and outcome emotions	• Situated framework • Relevant for achievement settings • Links emotions to motivation, cognition, self-regulation and achievement • Provides framework linking control/value combinations to emotion patterns	• Excludes other relevant emotions activated in education (e.g., social emotions) • Dichotomizes control, value and achievement variables • Does not fully address temporal nature of emotions • Universality of assumptions need to be tested across educational settings	• How are control and value appraisals related to emotions in med ed? • Do differences in control/value appraisals emerge across environment and experience level? • What differences in emotions emerge between activities and outcomes in med ed? • How do emotions relate to cognition, motivation, and performance?
Emotion regulation theories	• Distinguished from emotion generation in that the goal is to modulate or modify emotion • May involve increase/decrease,	• Not anchored to any specific emotion framework • Construct has utility in terms of predicting achievement and well-being	• Difficult to discern between emotion-generation and emotion-regulation • Existing taxonomies emphasize regulation of negative rather than positive	• What strategies are used to manage emotional states in med ed? • Is regulation automated or deliberate? • Is knowledge of these

	or maintenance of positive or negative emotions • Organized into different families of regulation strategies	• Provides more complete understanding of emotional experiences and goal-striving	emotions • Universality of regulation strategies need to be tested across educational settings	strategies explicit? • Which strategies are more adaptive? • Do regulation strategies differ across experience level and environment or task within med ed?

Table 3

Summary of review of emotion measures and applications for medical education

Data channel	Sample measures	Technology advances	Advantages	Disadvantages	Medical education settings
Self-report	• Questionnaires • Interviews • Experience sampling • Emote-alouds	• Wifi-enabled and tablet accessible for increased mobility, efficiency, and storing	• Economical • Efficient • Scalable • Captures subjective experience	• Difficult to capture continuous or dynamic nature of emotions • Susceptible to cognitive and memory biases	• Classrooms • Simulations (difficult to collect during performance tasks) • Naturalistic (difficult to collect during performance tasks) • CBLEs
Behavioral	• Facial expression • Speech and paralinguistics • Body language	• Automatic detection software for coding (facial recognition software; speech analyses software)	• Not susceptible to cognitive and memory biases • Provides continuous stream of process data	• Human coding is time intensive • Software is reliant on databases of posed emotions; not well-equipped to detect subtle, spontaneous emotions • Quality of video and audio data vary • Multi-modal interference • Susceptible to impression management	• Classrooms (ideal with individual or small group learning) • Simulations (ideal for body language) • Naturalistic (recording equipment can be intrusive in some settings) • CBLEs (ideal for facial)
Physiological	• Central nervous system (regional brain activation) • Autonomic nervous system (electrodermal activity, heart rate, salivary cortisol test)	• Brain imaging studies • Mobile physiological measures	• Provides continuous process data • Not susceptible to attempts to conceal emotion or cognitive and memory bias • Fine-grained data • Some measures are more mobile and discrete	• Only assess physiological component rather than valence or discrete emotions • Brain imaging studies are expensive, low ecological validity • Susceptible to interference by non-affective factors	• Classroom • Simulations • Naturalistc • CBLEs

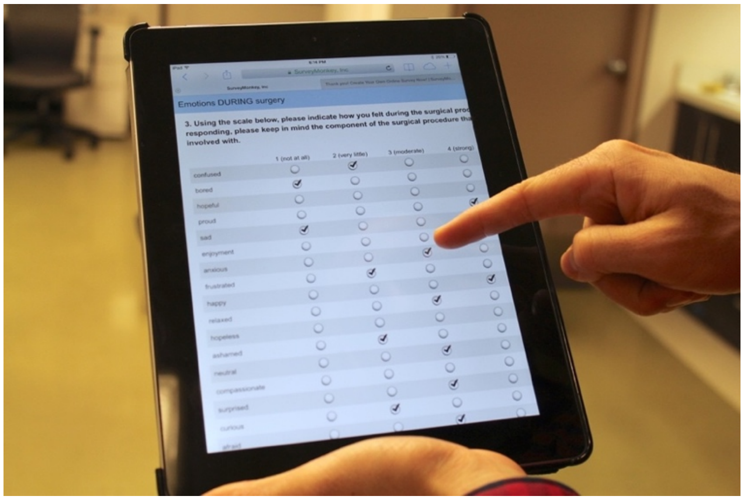


Figure 1. Self-reported concurrent emotional states using WiFi enabled tablet

Figure from “Measuring Emotions in Medical Education: Methodological and Technological Advances within Authentic Medical Learning Environments,” by M. C. Duffy, S. P. Lajoie, & K. Lachapelle. In S. Bridges, L. K. Chan, & C. E. Hmelo-Silver (Eds.), *Educational technologies in contextual learning: Research in health professions education* (p. 192), 2016, Switzerland: Springer: Copyright 2016 by Springer. Reprinted with permission.

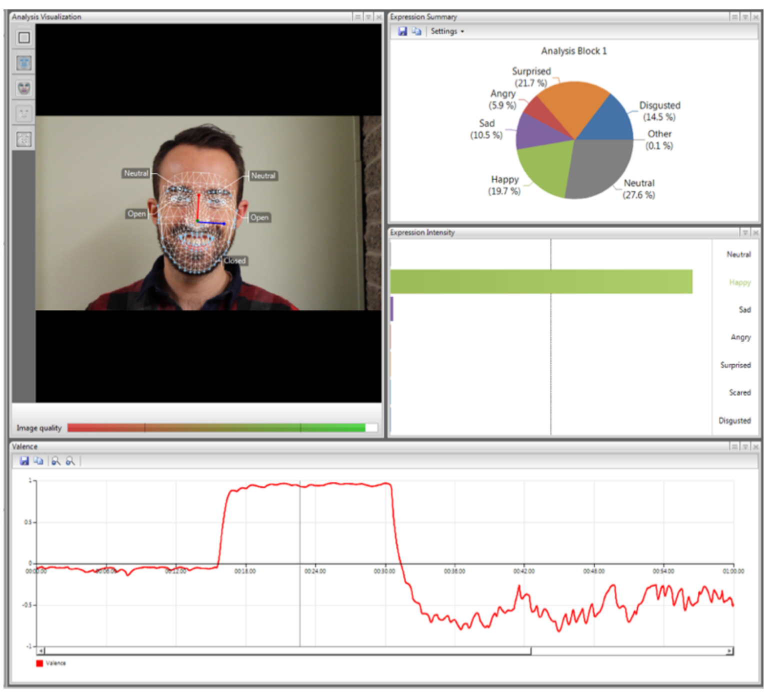


Figure 2. Automated emotion classification of facial expression within computer-based learning environment using video and FaceReader™ software

Figure from “Measuring Emotions in Medical Education: Methodological and Technological Advances within Authentic Medical Learning Environments,” by M. C. Duffy, S. P. Lajoie, & K. Lachapelle. In S. Bridges, L. K. Chan, & C. E. Hmelo-Silver (Eds.), *Educational technologies in contextual learning: Research in health professions education* (p. 198), 2016, Switzerland: Springer: Copyright 2016 by Springer. Reprinted with permission.

Figure 3. Affectiva Q™ sensor is used to track electrodermal activity in dynamic environments

Figure from "Measuring Emotions in Medical Education: Methodological and Technological Advances within Authentic Medical Learning Environments," by M. C. Duffy, S. P. Lajoie, & K. Lachapelle. In S. Bridges, L. K. Chan, & C. E. Hmelo-Silver (Eds.), *Educational technologies in contextual learning: Research in health professions education* (p. 201), 2016, Switzerland: Springer: Copyright 2016 by Springer. Reprinted with permission.

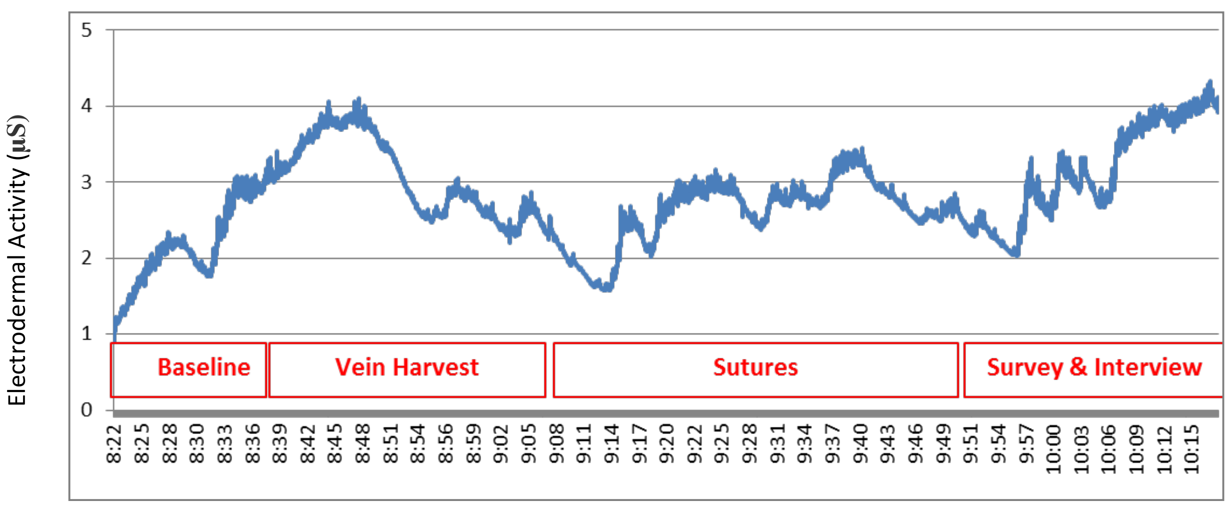


Figure 4. Electrodermal activity data over time within surgical environment

Figure from "Measuring Emotions in Medical Education: Methodological and Technological Advances within Authentic Medical Learning Environments," by M. C. Duffy, S. P. Lajoie, & K. Lachapelle. In S. Bridges, L. K. Chan, & C. E. Hmelo-Silver (Eds.), *Educational technologies in contextual learning: Research in health professions education* (p. 202), 2016, Switzerland: Springer: Copyright 2016 by Springer. Reprinted with permission.

Bridging Text

In Chapter 2, I presented a comprehensive review of theories and methods used to conceptualize and measure emotions. The goal was to consider the relevance and applications of different approaches for medical education and to identify ways in which to advance the field by critically reviewing previous research that has examined emotions among medical professionals. Based on the results of the review, it was concluded that relatively little research within medical education has measured emotional experiences beyond stress responses. Key areas for moving the field forward were identified, including: (1) examining change in positive and negative emotions over time, (2) comparing expert/novice emotions, and (3) measuring state emotions. In terms of theories, it was determined that control-value theory of achievement emotions could provide a useful framework for understanding the nature and role of emotions in medical education given its relevance for achievement activities. In terms of methodologies, it was concluded that using multiple methods to triangulate data and examine complementary lines of inquiry could lead to a richer understanding of the role of emotions in medical education.

In Chapter 3, I present an empirical investigation that follows from the recommendations and conclusions garnered from the comprehensive review. The study represents a situated approach to examining state emotions within medical education and was conducted within the domain of surgery, specifically the operating room. Multiple data channels and mixed methods analyses were used to examine the nature and role of emotions for an important surgical procedure within two authentic surgical training conditions: (1) a real case, and (2) a simulation. In doing so, the study seeks to explore the key areas identified in the comprehensive review to move the field forward using control-value theory as a framework to guide the constructs and hypotheses regarding appraisals, emotions, performance and task outcomes across experience level and training conditions.

Chapter 3: Manuscript 2

Real case operating room versus in situ simulation: Examining relations between appraisals, emotions, performance, and task outcomes across surgical training conditions

Earlier versions of this manuscript were presented at the following conferences:

Duffy, M. C., Lajoie, S., Pekrun, R., & Lachapelle, K. (2015, August). *Examining emotional states and technical skills performance in medical education*. Paper presented at the conference of the Junior Researchers of the European Association of Research on Learning and Instruction, Limassol, Cyprus.

Duffy, M. C., Lajoie, S., Pekrun, R., Ibrahim, M., & Lachapelle, K. (2015, March). *Simulation versus operating room: Examining relations between emotional states, learning processes, and technical skills performance across training environments*. Poster presented at the consortium of the American College of Surgeons Accredited Education Institutes, Chicago, IL.

Abstract

Within surgical education, emotions have received relatively little attention. In the broader education literature, however, emotions have been shown to play an important role in learning and performance. To address this gap, we examined relations between appraisals, emotions, performance, and task outcomes for a surgical procedure conducted during coronary artery bypass surgery in two operating room conditions: (1) a real case, and (2) a simulation. Questionnaires, interviews, video recordings, and physiological data were collected. Using a quasi-experimental design, 14 participants conducted a simulated vein harvest procedure and 19 participants conducted a real vein harvest procedure. The results from quantitative and qualitative analyses were largely consistent with our hypotheses. Correlation and mediation analyses revealed that control appraisals were negatively related to negative emotions (e.g., frustration, confusion, anxiety), which were negatively related to technical performance and task outcomes (confidence, satisfaction). Negative emotions, which were particularly detrimental for performance and task outcomes, were highest among novices. However, results showed reappraisal regulation strategies to play a useful role in managing these types of emotions and improving performance, particularly in the real case as compared to the simulation condition. Although learning gains may be optimized when emotions are less intense during initial training in simulation environments, enhancing the psychological fidelity of simulations and scaffolding effective regulation may help more experienced trainees become more emotionally prepared to effectively manage real cases. This study is the first to test and provide evidence supporting control-value theory within the surgical environment. Findings from this work improve our understanding of the role of emotion within authentic medical learning environments and can help inform the design of educational interventions.

Keywords: emotion; surgical education; technical performance; simulation

Introduction

What role do emotions play in surgical training and performance? What factors contribute to differences in emotional states in the operating room? Which regulation strategies are useful for managing emotions during surgical procedures? Technical skill development is the cornerstone of surgical education as performance within the operating room directly impacts patient outcomes. Although previous research has attended to the effectiveness and efficiency of diverse surgical techniques, it is important to understand how psychological processes, such as emotions, relate to technical performance for trainees and practitioners. Furthermore, emotions can have important implications for other training outcomes, such as confidence and motivation. Thus, a closer examination of emotions within surgical environments is needed in terms of its relations to appraisals, regulation, and performance, as well as variations in emotional profiles across experience levels and training conditions.

This study is the first of its kind to address this gap by examining emotional states during a surgical procedure for both real and simulated cases in the operating room. The present study advances emotion and medical education research by using a multi-method approach to examine the nature and functional role of emotions within an authentic medical learning environment; specifically, the surgical operating room. The following sections provide background on surgical training, technical performance, and the application of control-value theory of achievement emotions to surgical settings.

Surgical Training: Clinical Practice and Simulation

The traditional mode of surgical training conforms to an apprenticeship model that primarily involves immersing trainees in the operating room to develop surgical knowledge and skills through instruction, feedback, and practice (Dunnington, 1996; Kerr et al., 1999; Tsuda et al., 2009; Walter, 2006). This process involves trainees observing and assisting procedures until they can conduct these techniques with sufficient autonomy and automaticity such that they are capable of training junior trainees (Kairys et al., 2008, Kenton, 2006); a learning trajectory often referred to in surgical education as “see one, do one, teach one” (Kotsis & Chung, 2013, p. 1194). However, training within the operating room during live patient cases presents pragmatic challenges such as prolonged operating time, increased demands on faculty workload, and associated expenses (Babineau et al., 2004; Bridges et al., 1999). Moreover, recent duty hour restrictions limit residents’ exposure to the operating room (Antiel et al., 2012; Lindeman et al., 2014). In light of these challenges, there has been an increase in the use of simulations to augment clinical training and enhance performance (Tsuda et al., 2009) as part of a noticeable shift towards competency-based training (Buckley et al., 2014).

In the context of education, simulation is considered a training technique that approximates key characteristics of a real-world scenario or process (de Montbrun, & MacRae, 2012; Gaba, 2004; Issenberg et al., 1999). Simulations can vary across a number of features including: fidelity (degree to which it appears life-like or replicates clinical practice); location (conducted within a classroom, hospital site, or simulation center); systems or materials (e.g., computer-based environment, mannequin, cadaver); and purpose or learning goals (e.g., procedural skills, team communication). In surgical education, for instance, bench models (synthetic materials or inanimate animal tissues) are commonly used for technical skills training outside of the operating room (de Montbrun & MacRae, 2012). Regardless of the location or design, a key advantage of simulation training is that it provides opportunities for individuals to conduct procedures more autonomously without risk to patients (Aggarwal et al., 2010; Sachdeva et al., 2011). Initial evidence further shows simulation training to accelerate learning curves and improve clinical performance (e.g., Blum et al., 2004; Sedlack et al., 2004; Sturm et al., 2008)[6]. As such, there has been a concerted effort to integrate simulation into surgical training programs. For example, within cardiac surgery, trainees (i.e., residents and students) often participate in a series of simulations that provide exposure to common heart surgery procedures, such as replacing/repairing heart structures (e.g., aortic valve replacement) and connecting blood vessels (e.g., anastomosis, see Baker et al., 2012).

Despite the benefits of simulation and clinical training, the mechanisms underlying learning processes and performance are poorly understood. Previous research has largely focused on pre-post training comparisons without consideration for learner-centered variables or psychological processes. In addition, understanding the degree to which these processes are consistent across training conditions involves assessing not only to what extent simulation replicates real case procedures and materials but also relevant emotional processes. To maximize learning opportunities, it is critical to understand the nature of trainees' and professionals' psychological processes, including emotions and their regulation, and how these variables relate

[6] A recent meta-analysis tempers these claims by noting limitations in the measurement of surgical performance due to an exclusive focus on completion time as opposed to metrics for quality or safety (Buckley et al., 2014)

to performance on surgical tasks. In the following section, the role of emotions in surgical performance is more closely examined with a specific focus on technical performance[7].

Surgical Performance and Emotions

Proficient technical skills are an essential competency in many medical domains, but perhaps no place is it more integral to successful patient outcomes than within surgery (Aucar et al., 2005). Although competency in surgery is a multi-faceted construct (Arora et al., 2009), key components include: knowledge of surgical procedures and instruments; fluid movement in instrument handling; accurate execution of surgical techniques; and efficient completion of operative steps (Reznick et al., 1997). These types of technical skills have often been the focal point of surgical training (Walter, 2006)[8] and are included within the core competency of patient care as described by the Accreditation Council on Graduate Medical Education (Williams et al., 2004, Tsuda et al., 2009). Mastering these skills requires complex motor skills (e.g., manipulation and coordinator of muscles movements), knowledge of physiology, anatomy, and operative steps, as well as visual and haptic cue recognition (e.g., perceiving and interpreting visual and tactile information, see Rogers et al., 2006, Walter, 2006).

Failure to effectively master or execute these skills can place a strain on health care systems. In fact, the majority of adverse events in hospitals[9] involve surgical complications (Baker et al., 2004; de Vries et al., 2008; Vincent et al., 2001) that lead to increased hospital stays and expenses (Dimick et al., 2004). Although checklists have helped to reduce errors (Haynes et al., 2009; de Vries et al., 2010), the persistence of adverse events (Stahel et al., 2010) suggests there is room for improvement. A better understanding of factors that relate to technical performance may help to improve patient outcomes.

While factors such as deliberate practice, expert feedback, and authentic environments are considered critical in the acquisition of motor skills, the role of affective factors in technical performance has received less attention (Kneebone, 2005). In fact, within the surgical literature, skills in decision-making, communication, teamwork, leadership, and stress management have been broadly cast into categories of "non-technical skills" (e.g., Agha et al., 2015; Flin et al.,

---

[7] This review focuses on the domain of surgery. For a review of research examining emotion in the broader domain of medical education see Duffy et al. (2016).

[8] For more detail on other aspects of surgical competence (e.g., leadership) see Aorora et al., 2009; Sachdeva, 2002; Walter, 2006.

[9] These adverse events refer to errors or complications leading to temporary disability or permanent impairment, which are unrelated to the underlying disease (Baker et al., 2004)

2007; Yule et al., 2006a) or "human factors" (e.g., O'Connor et al., 2010; Shouhed et al., 2012). Of these factors, emotions has received far less attention than teamwork or communication[10], despite claims that management of affective states are important for effective surgical performance (Arora et al., 2010a; Hull et al., 2012; Yule et al., 2006a). For instance, measures developed to assess non-technical skills in surgery do not explicitly include emotion variables (e.g., Sharma et al., 2011) nor do articles that discuss instruction of non-technical skills incorporate strategies to manage emotional states (e.g., Flin et al., 2007). However, the results from Hull et al.'s (2012) systematic review provide compelling evidence that a significant (albeit complex) relationship exists between affective factors and technical performance, which they assert influences surgical outcomes.

Whereas some research has revealed that stress improves technical performance for surgical procedures (e.g., LeBlanc, 2008), other studies have found that it impairs performance (e.g., Arora et al., 2010b). In terms of factors impacting affective states, have been detected according to the nature of the task and demands. LeBlanc et al. (2008), for example, found that among surgical residents, stress was higher for a high-stakes simulation condition compared to low-stakes condition. Studies have also shown that stress is lower for open surgical procedures compared to laparoscopic procedures[11] (Berguer et al., 2001; Smith et al., 2001). Tendulkar et al. (2005) and Yamamoto (1999) found stress decreased with experience level. Qualitative research among surgical staff and trainees has corroborated these findings, with factors such as task complexity, team dynamics, and equipment challenges cited as factors that contribute to stress during surgery (Wetzel et al., 2006).

As evident from the preceding description of existing research, research conducted in the surgical domain (as in other areas of medical education) has focused exclusively on stress experiences rather than on a range of emotional states. There is a need to move beyond an exclusive focus on negative emotions given that both positive and negative states are likely to occur within surgical environments. To understand what factors give rise to different emotional states, and how various emotions can, in turn, influence performance and learning outcomes in surgical settings, I turn next to the control-value theory of achievement emotions.

---

[10] Within frameworks of non-technical skills (e.g., Yule et al., 2006a, 2006b), affective factors have been referred to as a *personal resource* that involves the ability to deal with stress and fatigue.

[11] Requires a unique surgical tool that requires additional training

Control-value Theory of Achievement Emotions

Control-value theory (CVT) of achievement emotions (Pekrun, 2006; Pekrun, Frenzel, Goetz, & Perry, 2007; Pekrun & Perry, 2014) has been widely used within educational and professional achievement settings (tasks in which success or failure is evaluated according to competency standards, Pekrun & Linnenbrink-Garcia, 2014) to examine the nature and function of emotions. This social-cognitive framework is particularly relevant for a surgical environment as an achievement-oriented domain with high-stakes outcomes, high task complexity, and high standards for success; an environment where intense emotions are likely to be elicited. According to this theory, variability in individuals' subjective perceptions of control (extent to which one feels personally able to successfully execute intended actions and outcomes) and value (extent to which one considers a task to be interesting, useful, or important) serve as proximal antecedents of the types of emotions activated before, during, and after an achievement task. For instance, if an activity is perceived to be controllable and is positively valued, then an individual is expected to experience enjoyment. In contrast, if an activity is perceived to be controllable but is negatively valued, anger is likely to occur (Pekrun, 2006; Pekrun et al., 2007). Once emotions are activated, they can influence task performance through cognitive (e.g., attention, working memory), motivational (e.g., persistence, effort), and regulatory processes (Pekrun et al., 2007; Pekrun & Perry, 2014). Corroborating this assumption, research has revealed that emotions such as anxiety, boredom, enjoyment, and hope impact individuals' effort, problem-solving strategies, and self-regulation that, in turn, influence performance and motivation-related task outcomes (e.g., Pekrun et al., 2002; Ranellucci, Hall, & Goetz, 2015).

These patterns are complex as they depend on the intensity and timing of the emotion activation, as well as the nature of the task and individual differences. Applying this to technical performance in a surgical context, if an individual is feeling relaxed during a procedure, this may help to reduce cognitive load and orient attention toward surgical sites that provide useful information about anatomical structures or handling unexpected complications. On the other hand, if an individual is feeling highly anxious, this may divert attention away from the task at hand and lead to a preoccupation with potential detrimental outcomes. Besides links to attention (e.g., Lindström & Bohlin, 2011; Yiend, 2010), emotions can impact other facets of cognition important for technical skill performance, such as memory (e.g., Wolf, 2009), reasoning (e.g., Blanchette & Richards, 2010; Quartz, 2009), and decision-making (e.g., Angie et al., 2011;

Bandyopadhyay et al., 2013; Starcke & Brande, 2012). Drawing from the broader literature of sport psychology and kinesiology, research has shown that emotions also impact motor movements (e.g., Fawver et al., 2012; Naugle et al., 2010, 2012; Stins & Beek, 2011). During a surgical procedure, for instance, emotions could impact motion flow, speed, and accuracy due to tremors, tension, or excessive force. According to CVT, emotions and appraisals are influenced by distal antecedents embedded in the environment (e.g., team support, performance culture, Pekrun & Perry, 2014). At the same time, individuals can deploy strategies to up/down regulate emotional states (Pekrun & Perry, 2014). In the following section, we delineate how the variables described in CVT inform the research objectives of the current study.

The Present Study

The goal of the present study was to examine emotional states among surgical trainees and experts during an authentic technical procedure within the operating room to understand how emotions relate to several facets of learning and achievement including task appraisals (i.e., control-value perceptions), processes (i.e., regulation strategies), performance (i.e., technical performance), and outcomes (i.e., self-efficacy, satisfaction, perceived skill development). Differences in these variables were examined according to experience level (i.e., expert versus novice) and task phases (i.e., emotions before, during, after procedure). These relations and mean differences were examined in both the real case and simulation condition to determine whether differences exist. Accordingly, our research questions (examined in both conditions) ask: (1) Do fluctuations in emotional states emerge across multiple points in time for a surgical procedure? (2) Do differences emerge across levels of surgical experience? (3) How do emotional states relate to task appraisals, processes, performance, and outcomes?

Using a quasi-experimental design, we used a matching method (Stuart, 2010) to achieve samples that were similar with respect to prior experience level (novices and experts) for the real case (42.10% novice) and simulation (42.86% novice). This was necessary given the observational nature of the study (i.e., examine emotions in a naturalistic setting) and practical constraints (e.g., clinical rotation and on-going training between conditions prevented random assignment or repeated-measures design). After surveying the environment, we chose to measure state emotions for a surgical procedure (saphenous vein harvest) during coronary artery bypass surgery. The rationale for selecting this procedure is explained in the methodology section.

In terms of hypotheses for correlational analyses, we expected that similar relations between emotions, task processes, and outcomes would occur across the two task conditions (simulation and real case). Based on control-value theory (CVT) of achievement emotions and the process model of emotion regulation (Pekrun, 2006; Gross, 2015), we predicted that appraisals of control and value for the task would be positively related to positive emotions, adaptive regulation strategies, performance, and task outcomes, but negatively related to negative emotions. We also expected that positive emotions and reappraisal strategies would be positively related to performance and task outcomes, whereas negative emotions would be negatively related to performance and task outcomes, and suppression strategies would be unrelated. Moreover, we expected the relationship between appraisals, performance, and task outcomes to be mediated by emotions.

In terms of hypotheses for expert/novice participants, we expected novices to report less control, more intense negative emotions, lower performance, and less positive task outcomes as compared to experts. We expected value and positive emotions to be similar among experts and novices in the real case since both groups would consider the task important and would elicit positive emotions (e.g., compassion, curiosity, enjoyment). We expected value to be higher for novices than experts in the simulation given the opportunity for greater skill improvement. In terms of differences between conditions, we expected the real case to elicit more intense emotions across participants at each time phase given the high-stakes nature of the task. In terms of change in emotions over time, we expected positive emotions to increase and negative emotions to decrease over time. We also expected novices to report more intense negative emotions over time compared to experts. We did not formulate specific hypothesis for the remaining variables given the paucity of research examining technical performance (Hull et al., 2012) for real operating room cases. The following section describes the measures and methods used to test our hypotheses.

Methods

Participants

Twenty-nine participants enrolled in this study. Sixteen medical residents and pre-residency medical students[12] (n = 3 female) on rotation in cardiac surgery participated in the real case condition. The mean year of residency was 2.19 (SD = 2.20). Medical students were in the

[12] Final year of medical school or equivalent on rotation in cardiac surgery.

final year or equivalent. Three cardiac surgeons with between 11-20 years of experience participated in the real case condition. Of the participants in this condition, 50% identified as Arab/Middle Eastern, 25% as Asian, 18.75% as Caucasian, and 6.25% as other/mixed ethnicity; the mean age was 30.31 (SD = 5.51). Twelve medical residents and pre-residency medical students (n = 5 female) on rotation in cardiac surgery participated in the simulation condition. The mean year of residency was 1.33 (SD = 1.72). Medical students were in the final year or equivalent. Two cardiac surgeons with 11-15 years of experience participated in the simulation condition. Of the participants in this conidtion, 43% identified as Arab/Middle Eastern, 7% as Asian, 50% as Caucasian; the mean age was 32.50 (SD = 8.46).

Measures & Apparatus

The complete list of scale items is presented in Appendix A.

Demographics questionnaire. Demographics were collected including age, gender, ethnicity and level of medical training/practice (e.g., program and year of education).

Prior experience scale. The prior experience scale included questions prompting participants to describe prior experience. It included a Likert scale to assess the number of times previously observing and conducting the procedure in the operating room and simulation environment. Option choices included: 0, 1-5, 6-10, 11-15, 16-20, 20+. Given that this question required participants to estimate their prior experience, option choices included a range of times (categorical), rather than requesting a fixed number that may be more susceptible to error. In addition, an open-ended question was included asking participants to report relevant training. Participants were classified as experts if they conducted the procedure more than five times[13].

Appraisals scale. A multi-item scale was used to measure subjective perceptions of control and value for the task. Scales were informed by expectancy-value theory (see Wigfield & Eccles, 2000) and tailored to the surgical procedure. In both conditions (real case and simulation), three items were included to assess participants' perception of control for the following aspects of the task and outcome: what they were expected to do for the procedure, their surgical performance, and surgical assistance. In the real case, an additional item was included to measure participants' perception of control over the patient's outcome (not

[13] This grouping represents relative rather than absolute expertise (Chi, 2006) and produced similar results to alternate grouping methods, such as median-split and composite experience scores (i.e., years of training, number of times observed and conducted procedure). This grouping allowed us to maintain sufficienct and similar sample sizes for novice and expert groups. With a larger sample, other experience levels (e.g., intermediates) can be examined.

applicable for the simulation). A sample item is: “Within this environment, I feel I have control over my surgical performance.” Three items were included to assess participants’ perception of value for the following aspects of the task: utility, interest, and importance (Eccles et al., 1983). A sample item is: “Conducting this surgical procedure is useful to me.” A 5-point Likert scale was used to rate agreement ranging from 1 (*strongly disagree*) to 5 (*strongly agree*).

Emotional state scale. The Medical Education Emotion Scale (MEES, Duffy et al., 2015) was used to measure 22 state emotions using single-item adjectives for each discrete emotion (e.g., anxious, proud, relieved, frustrated), which belong to four categories of emotion taxonomies: basic, epistemic, social, and achievement. These discrete emotions can also be grouped according to (1) valence: positive and negative[14], and (2) arousal level: activating versus deactivating. These can be combined to form four categories: positive activating, positive deactivating, negative activating, and negative deactivating emotions. Participants were instructed to keep the task in mind when indicating the intensity of their current emotion. Instructions were tailored to keep in mind the relevant time point; before, during, or after surgical procedure (e.g., “*keeping in mind the surgical procedure you are about to conduct, indicate how you currently feel. For each emotion, please indicate the strength of that emotion by selecting the number that best describes the intensity of that emotion*”). A 5-point Likert scale was used to rate intensity of each emotional state ranging from 1 (*not at all*) to 5 (*very strong*).

Emotion regulation scale. Gross and John’s (2003) 10-item scale was used to measure participants’ self-reported use of the following two emotion regulation strategies: cognitive reappraisal and emotion suppression. A sample item from the reappraisal sub-scale is: “When I wanted to feel more positive emotions, I changed the way I was thinking about the situation.” A sample item from the suppression sub-scale is: “When I was feeling negative emotions, I made sure not to express them.” A 7-point Likert scale was used to rate agreement that a given emotion regulation had occurred ranging from 1 (*strongly disagree*) to 7 (*strongly agree*).

Self-efficacy scale. A single item adapted from Bandura’s (2006) scale was used to measure participants’ confidence that they could effectively conduct the procedure. The item was phrased as follows: “Rate how confident you are that you can effectively conduct this surgical procedure,” A 10-point Likert scale was used to rate confidence level ranging from 0 (*no confidence*) to 100 (*complete confidence*) using 10-point intervals.

[14] Neutral and surprise were not included in analyses for the present study, as they are non-valenced states.

Proficiency scale. A single item was used to measure participants' self-assessed skill level in terms of conducting the surgical procedure. The item was phrased as follows: "Rate your level of proficiency in conducting this surgical procedure." A 5-point Likert scale was used to rate proficiency ranging from 1 (*poor)* to 5 (*very good*).

Performance satisfaction scale.[15]A single item was used to measure participants' satisfaction with their performance on the surgical procedure. The item was phrased as follows: "I am satisfied with my performance during the surgical procedure." A 5-point Likert scale was used to rate agreement that the participant was satisfied with their performance ranging from 1 (*strongly disagree*) to 5 (*strongly agree*).

Simulation model feedback scale.[16] A 2-item feedback scale was used to assess the realism and utility of the simulated vein model. The realism item was phrased as follows: "I found the simulator vein model realistic." The utility item was phrased as follows: "The vein model was useful for practicing my technique." A 5-point Likert scale was used to rate agreement ranging from 1 (*strongly disagree*) to 5 (*strongly agree*).

Technical skills performance. The Objective Structured Assessment of Technical Skill (OSATS, Martin et al., 1997; Reznick et al., 1997) global rating scale contains 7 dimensions that measure the multi-faceted nature of operative skill, including *respect for tissue, time and motion, instrument handling, knowledge of instruments, flow of operation, use of assistants,* and *knowledge of specific procedure*. For each item, anchored descriptions were provided for the lower end, midway point, and upper end of the 5-point Likert scale to operationally define skill level. For instance, the anchor for the lower end of the tissue handling item reads: "repeatedly makes tentative or awkward moves with instruments by inappropriate us of instruments"; whereas, the anchor for the upper end reads: "fluid moves with instruments and no awkwardness." The complete scale is presented in Appendix B. Although alternative rating scales have been developed to assess technical skill competencies and performance, the OSATS is appropriate when used for a specific procedure, such as the vein harvest, as opposed to an entire operation. In addition, high levels of reliability and validity have been established for this assessment tool (MacRae et al., 2000; Reznick et al., 1997).

---

[15] Other aspects of perceptions towards performance were also collected; these were not analyzed in the present study but can be used in future analyses to measure calibration between participant and expert performance ratings.
[16] After piloting an extended version of the vein model feedback scale, it was observed that participants provided more information via interview than questionnaire responses regarding their attitudes toward the vein model. As such, extended questions regarding the simulated model were posed via interview rather than questionnaire.

Eye-tracker device. SensoMotoric Instruments (SMI, Berlin, Germany) portable eye-tracking glasses version 1.0 (sampling frequency rate of 30 hertz) was used to capture scene recordings (video recordings that reveal participants' view of the environment/visual field) and eye gaze recordings (recordings of eye movements to measure fixations, saccades, and blinks). The eye-tracking glasses contain three cameras embedded within the frame of the glasses (two used to capture eye movements using infrared light projected toward the eyes and one used to capture scene recordings). Data were collected using the SMI iViewX™ 1.2 software. In the present study, eye-tracking recordings were used only for video review of task performance by expert raters (excludes eye-gaze behavior), which were replayed for post-task analyses.

Electrodermal activity sensor. The portable Affectiva Q™ sensor bracelet 2.0 (sampling frequency rate of 8 hertz) was used to measure participants' electrodermal activity level as an index of physiological arousal (Affectiva, 2014). More specifically, electrodermal activity is measured by placing the two electrodes embedded in the device directly against the skin. A small electrical current is emitted between the electrodes and allows for variations in electrical conductance on the skin to be measured based on the ease with which the electric current passes through the skins as the pores begin to fill with sweat (unit of measurement in microsiemens, μS). These changes in skin conductance are theoretically linked to sympathetic activity in the brain (e.g., stress), which is evidenced by changes in physiological arousal (e.g., increased sweating). Previous research has used this device across a variety of educational settings as it is unobtrusive and provides long-term, continuous data on physiological arousal (Azenberg & Picard, 2014; Kapoor, Burleson, & Picard, 2007; Picard, 2010; Pow, Swenson, & Picard, 2010; Woolf et al., 2009).

Interviews. Semi-structured interview questions were designed to complement questionnaire responses and online data traces by gaining insight into participants' perspective and by allowing participants to further elaborate on their experience in their own words. Questions and follow-up probes were used to address three key themes in a linear sequence as follows: (1) emotional states (positive, negative, and neutral), (2) factors that contributed to emotions (proximal and distal antecedents), and (3) regulation strategies used to manage their emotions. Participants were first prompted to reflect on the saphenous vein harvest procedure they had just completed and then invited to draw comparisons to other surgical procedures and training environments. A mix of open-ended and structured questions were asked. As a validity check of the self-report questionnaire, participants were asked after the real case whether the emotion questionnaire items seemed appropriate and whether any relevant emotional states were lacking from the list of adjectives. As a validity check of the simulation model, participants were asked questions after the simulation regarding the fidelity and utility of the vein model. Interview questions for the real case and simulation are presented in Appendix B and C.

Surgical Task

According to the World Health Organization (2014), coronary artery disease (reduced blood supply to the heart caused by plaque accumulation) is the leading cause of death worldwide. In 2013 it was responsible for over 8 million deaths (GBD Collaborators, 2015). Coronary artery graft bypass (CABG) surgery remains the standard of practice to treat severe cases (Mohr et al., 2013, p. 629; Farkouh et al., 2012; Gulati et al., 2009) and is one of the most commonly conducted operations (American College of Cardiology and American Heart Association Task Forces, 2004). The surgery involves harvesting a graft from a viable vein or artery elsewhere in the body to bypass blocked arteries in the heart and restore normal blood flow (National Heart, Lung, and Blood Institute, 2012). One graft commonly used is the greater (or long) saphenous vein of the leg (Hess et al., 2014; Zenati et al., 2011).

We surveyed multiple technical procedures (e.g., consulted with subject matter experts, observed in operating room) to select an appropriate task for this study that was commonly used for training surgical skills, as well as sufficiently important and complex to include participants with a range of experience (i.e., novice and experts). We chose the saphenous vein harvest procedure given that it is a core component of surgical training for novices, as well as a common procedure for more experienced cardiac residents and surgeons. In a conventional open vein harvest procedure (see Doty & Doty, 2012), the first step is to identify the location of the vein and make an incision. The incision may extend from the ankle to below the knee or mid-thigh, depending on the length of graft needed. Subsequent steps involve exposing and freeing the vein by dissecting and removing connective tissue and by clipping and tying side branches. Once desired length is obtained, the vein is clamped and tied at one end and then distended by inserting a cannula (tube) and injecting a solution to preserve the vein and identify any perforations that require repair. The incision site is then closed using sutures. Figure 5 provides an illustration of this technical procedure.

Conditions

Using a quasi-experimental design, participants completed one of two conditions: (1) real case or (2) simulation. Both conditions took place within the operating room at a teaching hospital affiliated with a major North American university. Data collection for the real case took place first over approximately 8 months. After this point, the simulation models were developed and constructed (the models were ideal to use shortly after construction to preserve the quality of materials). Data collection then began for the simulation case and took place over approximately 7 months. Separate participants were recruited for these conditions, with the exception of 4 participants who participated in both the real case and simulation case. These participants were also included as experts in the simulation given that this level of expertise is acquired by a highly specialized group of individuals. Given the time between participation in the two conditions (approximately 1 year apart), the risk of confound was minimal. Each condition is described below.

Real case. Participants conducted a saphenous vein harvest procedure within a live operating room scheduled for a coronary artery bypass surgery for a real patient. Supervision and direction for the task was provided to varying degrees according to the participant's skill and experience level. For instance, novices were supervised by a senior resident or staff surgeon, were closely directed, and received explanation, demonstration, and feedback for surgical techniques, whereas staff surgeons conducted the procedure without supervision. In all cases, the surgical team was present as they conducted the procedure. This procedure provided medical trainees with an opportunity to improve technical skills and learn about surgical procedures in a naturalistic setting.

Simulation. The simulation vein models were developed at a medical simulation centre by a cardiac resident and surgical skills coordinator. The vein model consisted of synthetic gel-filled saphenous vein and branches, as well as synthetic connective skin and fat tissues. The model was attached to the leg of a mannequin (simulated patient), which was placed on an operating table and draped. Surgical instruments were available. Participants were instructed that they would be conducting the vein harvest procedure using the simulated vein model (one-time use as the model could not be re-used once dissected and sutured). The model construction was described and participants were informed that they could receive surgical assistance and instruction as necessary. Participants conducted a saphenous vein harvest procedure within an operating room that was not currently in use for surgeries but contained all equipment and instruments (in situ simulation). A cardiac resident was present to provide surgical assistance (e.g., passing instruments) or instructional aid as needed. Similar to the real case, supervision and direction for the task was provided to varying degrees according to the participant's skill and experience level. This procedure provided medical trainees with an opportunity to improve technical skills and learn about surgical procedures within an in situ simulation.

Experimental Procedure

The experimental protocol was largely the same for the two conditions (exceptions are noted in the following description). A staff surgeon and cardiac resident (co-authors of this study) identified potential participants currently on rotation in the cardiac unit, as well as dates and times for the scheduled surgery during which the surgical procedure would take place. After meeting the potential participant, the experimenter explained the purpose and protocol of the study, described what the task and surgical procedure would entail, provided the consent form for review, and invited questions. Participation was voluntary and participants could withdraw participation at any time. Once consent was provided, the Q sensor bracelet was administered. For the real case, the bracelet was administered on the lower leg/above the ankle as the bracelet could not be worn on the wrist during hand-washing before surgery. For the simulated case, two Q sensor bracelets were worn (one around the lower leg and one around the inner wrist) to determine whether there was consistency in measurements when applied to different sites of the body. In both conditions (real case and simulation), baseline measures of electrodermal activity were collected for at least 10 minutes prior to beginning the task when possible. During this time, the participant completed a series of questionnaires outside of the operating room, which included (in the following order): the demographics questionnaire, prior experience scale, appraisals scale, self-efficacy scale, proficiency scale, and pre-task emotional states scale. Participants were instructed that they would be completing a scale for the same list of emotions before and after the task and that they would be asked to reflect back on emotions they felt during the task, which would provide a total of three time points for emotional states in reference to the task: (1) before (measured concurrently), (2) during (measured retrospectively), and (3) after (measured concurrently). Given that we could not interrupt the procedure in the real case to administer the emotions scale concurrently, we opted to follow the same procedure in the simulation (participants provided retrospective accounts of emotional states experienced during the task) to limit possible confounds. This method of administration has been used in previous research (Arora et al., 2010b). Instructions for questionnaire items were tailored to the surgical task participants were conducting and key time points (e.g., "keeping in mind the surgical procedure you are about to conduct/just completed…"). Upon completion of the questionnaire, the eye-tracking device was fitted to the participant and calibrated. During the surgical procedure, eye-tracking and electrodermal activity data was collected. The experimenter was also present in the operating room and took notes and time-stamps for key phases of the surgical procedure and degree of assistance required. Once the surgical procedure was complete, the eye-tracker was removed and the participant and experimenter relocated outside the operating room. The participant completed a series of questionnaires, which included (in the following order): the during-task emotional states scale, post-task emotional states scale, self-efficacy scale, proficiency scale, satisfaction with performance scale, and emotion regulation scale. After the simulation case, participants also completed the vein model feedback scale. Following completion of these questionnaires, an interview was conducted. A tablet was used to access questionnaires and to record interviews. At the end of the interview, the Q sensor was turned off and removed. Figures 6 and 7 provide an illustration of the experimental set-up.

Coding and Data Processing

Technical skills performance data. Two raters (a surgeon and a resident) independently assessed videos obtained from eye-tracking recordings (i.e., scene view recordings) to assess technical skill performance for each participant using the Objective Structured Assessment of Technical Skill (OSATS, Martin et al., 1997) global rating scale. Raters were blind to experience level and identification of participants. To establish consistency between raters, an expert, intermediate, and novice video was selected beforehand for both the real case and simulation. Raters viewed and independently assessed technical skills for these participants before moving forward with the remaining sample. Discrepancies in these initial ratings were identified and resolved by discussing the rationale for decision-making. It was agreed that the upper range would represent expert performance whereas the lower range would represent novices. The midway point represented standard vein extraction.

Samples of key components of the surgical procedure (e.g., initial incision, dissection/extraction, tying branches, cannulation, sutures, knot-tying) were viewed by each rater until they were satisfied they had reviewed sufficient video content to make an informed assessment. The use of video recordings to assess performance post-hoc has been found to be both efficient and highly reliable in previous research (Dath et al., 2004). In particular, the use of global rating scales, such as the OSATS, for video-based assessments of surgical performance has been shown to distinguish experts from novices (Aggarwal et al., 2008).

Mean scores of each participant were calculated using the average of raters' total score (across rating scale items) for each participant. The mean scores were used as the technical skills performance variable in analyses. Inter-item consistency of the global rating scale was high for the real case ($\alpha = .97$) and simulation ($\alpha = .96$). Inter-rater reliability was conducted using an intra-class correlation[17], which revealed a high level of reliability between the two raters for the real case, ICC = .95) and simulation (ICC = .93).

Eye-tracking video data. Eye-tracking videos were reviewed and segmented using time-stamps to identify components of the procedure: (1) vein extraction (initial skin incision, dissection/vein exposure, inserting cannula and syringe, tying off branches or clipping them), and, (2) suturing (closing wound). It should be noted that not all steps are linear; in some cases, this is a cyclical process. For the real case, mean duration of vein extraction was 31.9 minutes (*SD* = 13.91) and 14.48 minutes (*SD* = 5.48) for suturing. For the simulation, mean duration of vein extract was 13.14 minutes (*SD* = 5.56) and 6.76 minutes for suturing (*SD* = 3.31).

Electrodermal activity data. Electrodermal activity log files were imported to Q sensor software and exported to excel files. Electrodermal activity data was segmented using time stamps from field notes for three key experimental and surgical events: (1) baseline/pre-task questionnaires; (2) vein harvest; (3) suturing; (4) post-task questionnaire/interview. The maximum value (i.e., highest skin conductance response level across all time points) was used for analyses. This value was used given that other descriptive statistics (e.g., mean) were more susceptible to recording errors such as loss of contact with skin or lack of activation during baseline. Correlation between the two device sites worn simultaneously (wrist and ankle) demonstrated high consistency ($r = .90$)

Interview data. The mean interview duration time was 9.38 minutes (*SD* =3.31) following the real case and 6.27 minutes (*SD* =2.52) following the simulation case. Following data collection, interviews were transcribed verbatim and coded by a research assistant trained by

---

[17] Two-way random consistency average measures (consistent sample of raters across sample, using mean for analyses (Landers, 2015)

the experimenter using the coding scheme presented in Appendix E. Piloting of the coding scheme led to minor revisions in an iterative fashion to further refine and tailor operational definitions to the context. A second coder (experimenter) verified final codes for consistency. Discrepancies were resolved through discussion, consultation with the coding scheme, and definitions of constructs according to the literature.

Results[18]

Validity of Training Conditions

Self-reported self-efficacy and proficiency ratings were examined pre-post within both conditions to assess whether these environments represented valid training opportunities by helping to improve task confidence and/or skill. We also examined alignment of performance ratings and prior experience (expert versus novice) as an index of task performance validity. For the simulation, we also examined the extent to which participants considered the simulated surgical model to be realistic and useful for practicing their vein harvest technique.

Real case. Participants reported stable task proficiency from pre-task (M = 3.74 out of 5, SD = 1.41) to post-task (M = 3.74, SD = 1.59) in the OR, as well a 1% increase in task self-efficacy from pre-task (M = 73.69 out of 100, SD = 28.52) to post-task (M = 74.73, SD = 26.42). Analyses of performance ratings revealed that the task could be used as a reliable indicator of performance, as raters were able to discriminate between experts and novices (84.21% accuracy between high/low global rating score and expert/novice experience level). Table 4 illustrates alignment between performance ratings and experience level for the real case.

Simulation. Participants reported a 7% increase in task proficiency from pre-task (M = 3.71 out of 5, SD = 1.69) to post-task (M = 4.07, SD=1.20) using the simulator, as well an 8% increased in task self-efficacy from pre-task (M = 74.29 out of 100, SD = 32.98) to post-task (M = 82.14, SD = 26.07). Analyses of performance ratings revealed that the simulator task could be used as a reliable indicator of performance, as raters were able to discriminate between experts and novices (92.86% accuracy between high/low global rating score and expert/novice experience level). Participants considered the simulation model to be realistic (M = 3.92 out of 5, SD = .83) and useful for practicing their vein harvesting technique (M = 4.15 out of 5, SD = .80). Table 5 illustrates alignment between performance ratings and experience level for the

[18] Analyses were conducted for the real case and simulation condition separately. Given that 4 participants completed both conditions and inferential tests require either independence of observations *or* repeated-measures for all participants, we opted for descriptive comparisons between conditions.

simulation. Qualitative analyses of interviews revealed that most participants found the model to be realistic in terms of anatomical accuracy, tissues, and procedure. The vein model was considered to be most beneficial for novices (i.e., junior residents, medical students) to learn procedural steps of the vein harvest, basic surgical skills (e.g., instrument handling, suturing), and build confidence to prepare for real OR cases. Several participants also noted it could be used by more experienced participants to practice or improve technique. More detailed feedback on the vein model and suggestions for improvement are provided in the qualitative component of the results section below.

Emotions over Time and Experience Level

Emotions over time (discrete categories): Real case and simulation. Descriptive statistics of the most intensely reported discrete emotional states before, during, and after the task are presented in Table 6 (real case) and Table 7 (simulation).

Emotions over time (valence and arousal categories). One-way repeated-measures Analyses of Variance (ANOVAs)[19] were conducted with experience level as the between-subjects factor (expert, novice) and time as the within-subjects factor (before, during, after task) to test for main effects of experience level and time, as well as interaction effects on positive activating, positive deactivating, negative activating, and negative deactivating emotions[20].

Real case. Results revealed a significant main effect of experience level on negative activating emotions, $F(1, 17)$, 14.34, $p < .01$, $\eta_p^2 = .46$, which were significantly more intense among novices than experts. There was a significant main effect of time on negative activating emotions, $F(2, 34) = 4.05$, $p < .05$, $\eta_p^2 = .19$, which pairwise comparisons revealed significantly decreased in intensity from during to after the task ($p < .05$). There was also a significant main effect of time on positive deactivating emotions, $F(2, 34) = 5.14$, $p < .05$, $\eta_p^2 = .23$, which pairwise comparisons revealed significantly increased from before to after the task ($p < .05$) and during to after the task ($p < .05$). There was a significant interaction effect on negative deactivating emotions, $F(2, 34) = 3.78$, $p < .05$, $\eta_p^2 = .18$, which pairwise comparisons revealed

[19] Multivariate analyses were not conducted as emotions with different arousal levels (activating vs deactivating) can have different patterns of effects; as such, combining multiple dependent variables into a composite score may obscure unique trends.
[20] In the simulation environment, negative deactivating emotions consisted of hopelessness and disappointment (excluded boredom, sadness), positive deactivating emotion consisted of relief (excluded relaxation) to ensure sufficient scale reliability as described in more detail in Table 11.

were significantly more intense for novices than experts after the task ($p < .05$). No other effects were statistically significant.

Simulation. Results revealed a significant main effect of experience level on negative activating emotions, $F(1, 12) = 6.87$, $p < .05$, $\eta_p^2 = .36$, which were significantly more intense among novices than experts. There was also a trend toward significance for experience on positive activating emotions, $F(1, 12) = 3.80$, $p = .075$, which were more intense among novices than experts. There was also a trend toward significance for experience on negative deactivating emotions (hopeless, disappointment), $F(1, 12) = 4.04$, $p = .068$, which were more intense among novices than experts. There was a significant main effect of time on negative activating emotions, $F(2, 24) = 5.76$, $p < .01$, $\eta_p^2 = .32$, which pairwise comparisons revealed significantly decreased in intensity from before to after the task ($p < .01$) and during to after the task ($p < .01$). There was also a significant effect of time on positive deactivating emotion (relief), $F(2, 24) = 5.37$, $p < .05$, $\eta_p^2 = .31$, which pairwise comparisons revealed significantly increased in intensity from before to after the task ($p < .05$) and from during to after the task ($p < .05$). There was also a trend toward a significant interaction effect on negative activating emotions, $F(2, 24) = 3.08$, $p = .064$, which showed the greatest difference between experts and novices during the task (more intense among novices). No other effects were statistically significant.

Relations between Task Appraisals, Emotions, Performance, and Task Outcomes

Bivariate correlations[21] were conducted to examine relations between appraisals (control, value), emotions (positive activating, positive deactivating, negative activating, negative deactivating), regulation (reappraisal, suppression), performance (technical performance scores, task completion time[22]), and task outcomes (post-task self-efficacy, satisfaction with performance, reported proficiency gains). Partial correlations were then conducted on these same relations to control for prior experience[23] based on the number of times the procedure was previously conducted. Correlations are presented in Table 8 (real case) and Table 9 (simulation). In the following section results from bivariate correlations are described followed by partial correlations for the real case and simulation.

[21] One-tailed tests were conducted as we had directional hypotheses (aprori) for relationships.
[22] Task completion time does not necessarily equate to quality of performance but as it is a component of how a task is performed, we include it in this grouping.
[23] Two-tailed tests were conducted as these analyses were more exploratory and we did not have specific hypothesis.

Real case.

Appraisals' relations to emotions, regulation, performance, and task outcomes. Results revealed that control was positively related to positive activating emotions (r = .56, p < .01), positive deactivating emotions (r = .61, p < .01) and negatively related to negative activating emotions (r = -.57, p < .01), negative deactivating emotions (r = -.55, p < .05). Control was also positively related to technical performance (r = .52, p < .05), satisfaction with performance (r = .66, p < .01), and post-task self-efficacy (r = .74, p < .01). Control also showed a trend toward significance with reappraisal regulation strategies (r = .37, p < .06). Although positively related to control appraisals (r = .39, p < .05), value was only significantly related to positively activating emotions (r = .44, p < .05) and was not significantly related to other emotions, technical performance, or task outcomes.

Emotions relations to performance, and task outcomes. Positive deactivating emotions were positively related to satisfaction with performance (r = .54, p < .01), proficiency gain (r = .40, p < .01), and post-task self-efficacy (r = .41, p < .05). Negative activating emotions were positively related to task completion time (r = .53, p < .05) and negatively related to technical performance (r = -.43, p < .05), satisfaction with performance (r = -.89, p < .01), proficiency gains (r = -.67, p < .01), and post-task self-efficacy (r = -.80, p < .01). Negative deactivating emotions were negatively related to satisfaction with performance (r = -.67, p < .01), proficiency gain (r = -.57, p < .01), and post-task self-efficacy (r = -.58, p < .01). Positive activating emotions were not significantly related to performance or task outcome variables.

Regulation relations to performance, and task outcomes. Reappraisal regulation strategies were positively related to technical performance (r = .39, p < .05) and post-task self-efficacy (r = .42, p < .05). Reappraisal strategies were also negatively related to physiological arousal as measured by electrodermal activity (r = -.48, p < .05). Suppression regulation strategies were not significantly related to performance or task outcomes.

Controlling for prior experience. Partial correlations revealed that control remained significantly related to positive activating emotions (r = .68, p < .01), positive deactivating emotions (r = .67, p < .01), and negative deactivating emotions (r = -.53, p < .03) but no longer was significantly related to negative activating emotions. In addition, control remained significantly related to post-task self-efficacy (r = .69, p < .01), and now satisfaction as well (r = .57, p < .01). In contrast to the previous analyses, value was significantly related to technical performance ($r = .55, p < .05$) when controlling for prior experience. In terms of emotions relations to performance and learning outcomes, negative activating emotions were now negatively related to task completion time ($r = -.51, p < .05$), post-task self-efficacy ($r = -.61, p < .01$), proficiency gain ($r = -.64, p < .01$) and performance satisfaction ($r = -.84, p < .01$). Positive deactivating emotions were positively related to performance satisfaction ($r = .60, p < .05$). Negative deactivating emotions were negatively related to post-task self-efficacy ($r = -.77, p < .01$), proficiency gain ($r = .55, p < .05$), performance satisfaction ($r = -.80, p < .01$).

Simulation.

Appraisals' relations to emotions, regulation, performance, and task outcomes. Control was positively related to technical performance ($r = .50, p < .05$). Value was significantly related to reappraisal strategies ($r = .58, p < .05$). No other relations were statistically significant.

Emotions relations to performance and task outcomes. Positive activating emotions were positively related to task completion time ($r = .59, p < .05$). Negative activating emotions were positively related to task completion time ($r = .62, p < .01$) and negatively related to technical performance ($r = -.73, p < .01$), satisfaction with performance ($r = -.69, p < .01$), and post-task self-efficacy ($r = -.91, p < .01$). Negative deactivating emotions were negatively related to technical performance ($r = -.60, p < .01$), satisfaction with performance ($r = -.79, p < .01$), proficiency gain ($r = -.55, p < .05$), and post-task self-efficacy ($r = -.59, p < .01$).

Regulation relations to performance and task outcomes. Regulation strategies were not significantly related to performance or task outcomes.

Controlling for prior experience. Partial correlations revealed that control did not show any significant relations to emotions, performance, or task outcomes when controlling for prior experience. Value was positively related to negative deactivating emotions ($r = .64, p < .05$). Negative deactivating emotions were negatively related to satisfaction with performance ($r = -.70, p < .01$). No other relations were statistically significant.

Mediation Path Analysis

To examine whether the relationship between appraisals and performance was mediated by emotions, a path analysis was conducted to test a mediation model using bootstrapping.[24] Path analysis was conducted only for the real case given the larger sample size and the fact that the

[24] Bootstrapping has been found to be a particularly useful technique for small sample sizes (Fritz & Mackinnon, 2007; Preacher & Hayes, 2008); there is evidence that it is robust with sample sizes as small as $N = 20$ (Creedon & Hayes, 2015).

relations between control, emotions, performance and task outcomes were all significant in this condition. Following Greene and colleagues' (2013, 2015) approach to select variables with the strongest bivariate relations to learning gains, the appraisal and discrete emotions that were most highly correlated with performance and post-task self-efficacy were selected for inclusion in the model (see Appendix F for a description of significant relations between discrete emotions, performance, and self-efficacy). We included emotions before and during the task to maintain the temporal nature of this model (emotions predict performance/task outcomes). The final model (Figure 8) included control, fear before the task, confusion during the task, performance, and post-task self-efficacy.

To test the hypothesized mediation model, we used Hayes and Preacher's (2014) PROCESS SPSS macro, which is recommended with complex mediational models as it maintains higher levels of power while controlling for Type I errors (Preacher & Hayes, 2008). We first calculated total effects models for the two outcome variables: performance and self-efficacy. The total effects model expresses the sum of the direct and indirect effects of control on these outcome variables. This step is conducted to determine the predictive relations between control independent of the effects of mediational variables. Following this, we calculated the direct effects of control predicting emotions; specifically, fear before the task and confusion during the task. Finally, we calculated the direct effects of emotions on the outcome variables, followed by the indirect effects of control on outcome variables via emotions, to determine if emotions significantly mediated relations between control and outcome variables.

Total effects of control on outcome variables. To test the hypothesis that control would positively predict outcome variables, two regression analyses were conducted separately predicting performance and self-efficacy scores. First, for the model predicting performance, control was entered as the independent variable. The overall model was significant, $F(1, 17) = 6.33$, $p < .05$, adj. $R^2 = .229$, indicating the predictor variable accounted for 22.9% of the variance in performance. Control was a significant positive predictor, $\beta = .52$, $t = 2.52$, $p < .05$. For the model predicting self-efficacy, the overall model was significant, $F(1, 17) = 21.03$, $p < .001$, adj. $R^2 = .527$, indicating the predictor variable accounted for 52.7% of variance in self-efficacy. Control was again a significant positive predictor $\beta = .74$, $t = 4.59$, $p < .001$

Emotions as mediators between control and outcomes. To examine the direct and indirect predictive relations between control, emotions, and outcome variables, parallel mediation analyses were conducted with PROCESS (Hayes & Preacher, 2014). This allows the estimation of all direct predictive effects of control and emotions simultaneously on the outcome variables (i.e., performance and self-efficacy). The following are two sets of regression analyses to test the hypothesis that control predicts emotions and two mediation analyses that include performance and self-efficacy to test the hypothesis that the relationship between control and outcomes is mediated by emotions.

Control was entered into a model that predicted fear. This model was significant, $F(1, 17) = 10.39$, $p = .005$, $R^2 = .379$, and accounted for 37.9% of the variance in fear. Control was a significant predictor, $\beta = -.62$, $t = -3.22$, $p = .005$, indicating that those with more control reported less fear before the task. The model predicting confusion was also significant, $F(1, 17) = 6.59$, $p = .02$, $R^2 = .279$, and accounted for 27.9% of the variance in confusion. Control was again a significant negative predictor, $\beta = -.53$, $t = -2.57$, $p = .02$, such that those with more control reported less confusion during the task.

Finally, to test the hypothesis that emotions mediate relations between control and outcome variables, we calculated full mediation analyses. Control was entered into the model predicting performance with fear and confusion. The overall model was significant, $F(3, 15) = 6.96$, $p = .004$, $R^2 = .582$, and accounted for 58.2% of the variance in performance. Compared to the total effects model, the inclusion of emotions as mediators resulted in an improved prediction of performance, $\Delta F(2, 15) = 5.57$, $p = .015$, $\Delta R^2 = .311$; however, control was no longer a significant predictor, $p > .25$. Confusion was a significant negative predictor of performance, $\beta = -.87$, $t = 2.50$, $p < .05$, indicating that more confusion predicted lower performance. Fear was not a significant predictor of performance $p > .25$. A test of the mediation paths showed that control had a positive indirect effect on performance that was significantly mediated by confusion, with a standardized point estimate of .46 and bias corrected bootstrapped confidence intervals (95%) of .17 to 1.10.

For the model predicting self-efficacy, the overall model was significant, $F(3, 15) = 17.92$, $p < .001$, $R^2 = .782$, and accounted for 78.2% of the variance in self-efficacy. Compared to the total effects model, the inclusion of emotions as mediators resulted in a significantly improved prediction of self-efficacy, $\Delta F(2, 15) = 7.87$, $p = .005$, $\Delta R^2 = .229$. Control remained a significant positive predictor of self-efficacy, $\beta = .37$, $t = 2.42$, $p < .05$. Fear was a negative predictor of self-efficacy, $\beta = -.57$, $t = -2.09$, $p = .054$. A test of the mediation paths showed that control had a positive indirect effect on self-efficacy that was mediated by fear (approaching significance), with a standardized point estimate of .35 and bias corrected bootstrapped confidence intervals (95%) of -.04 to .80. Confusion during surgery was not a significant mediator for self-efficacy.

Difference in Variables across Experience Level

A series of independent samples t-tests were conducted to examine whether experts and novices (between-subjects factor) differed in terms of: appraisals (control, value), emotions (positive activating, positive deactivating, negative activating, negative deactivating), regulation (reappraisal, suppression), technical performance, and post-task self-efficacy. Results for real case and simulation are presented below. Descriptive statistics for experts and novices are presented in Table 10 (real case) and Table 11 (simulation).

Real case. Results revealed a significant difference in terms of control, $t(17) = -3.05$, $p < .01$ (novices reporting less control), negative activating emotions, $t(17) = 3.35$, $p < .01$ (novices reporting greater intensity), technical performance, $t(17) = -6.00$, $p < .001$ (experts demonstrating higher performance), and post-task self-efficacy $t(17) = -4.70$, $p < .01$ (experts reporting greater self-efficacy). As three variables did not meet the assumption for homogeneity of variance, the Welch/Brown-Forsythe test statistics and p-values were used, but this did not affect significance. No other variables demonstrated statistically significant differences between experts and novices.

Simulation. Results revealed a significant difference in terms of control, $t(12) = -2.94$, $p < .05$ (novices reporting less control), negative activating emotions, $t(12) = 2.62$, $p < .05$ (novices reporting greater intensity), negative deactivating emotions, $t(12) = 2.31$, $p < .05$ (novices reporting greater intensity), technical performance, $t(12) = -4.16$, $p < .01$ (experts demonstrating higher performance), and post-task self-efficacy $t(12) = -2.95$, $p < .05$ (experts reporting greater self-efficacy). As two variables did not meet the assumption for homogeneity of variance, the Welch/Brown-Forsythe test statistics and p-values were used, but this did not affect significance. No other variables demonstrated statistically significant differences between experts and novices.

Qualitative Analyses

Content analyses of interviews were conducted to identify key patterns and categories pertaining to three themes: (1) emotions experienced during the simulation and real case; (2) antecedents for these emotions, and (3) regulation strategies. Additional themes emerged, such as comparisons between experiences of real case and simulation training, as well as optimal emotional state. Within the simulation environment, we also analyzed responses regarding the fidelity and utility of the simulation model. Categories were generated from interview responses through an iterative process of developing the coding scheme until saturation was reached. The results of analyses (including specific categories and frequencies of responses, corresponding sample quotations, and key patterns across participants) are presented for each theme in Table 12 for the real case and Table 13 for the simulation. Given that there is a great deal of consistency in participants' responses across conditions, these results are integrated in the section below.

With respect to emotions experienced, analyses of interviews from both the real case and simulation demonstrated that participants reported emotions consistent with questionnaire responses and multiple emotional states were often described. Commonly reported positive emotions for both the real case and simulation included *happiness/enjoyment* and *relaxation/calmness*. Other positive emotional states included *relief* and *gratitude*. In terms of differences across conditions, *pride* was more commonly reported in the real case, whereas *curiosity* was more commonly reported in the simulation. *Anxiety* was the most commonly reported negative emotion across both conditions. Other commonly reported negative emotions included states of *frustration* and *confusion*. Other negative states included *hopelessness* and *disappointment*. *Neutral* and *surprise* were also reported (as non-valenced states). The only affective state that appeared to be more blended or complex (i.e., difficult to link to a specific discrete state) was *confidence* which was often described as a feeling state related to combinations of discrete emotions, such as calmness, pride, and enjoyment. *Disgust* was not reported as an experienced emotion and in follow-prompts several participants suggested that it was not relevant for this context. *Boredom* was not reported as an intensely experienced emotion. *Excitement* was often reported and could be included as a scale item in future work. The object-focus of these emotions (i.e., what the emotions were directed towards) included the task (challenges, problem-solving, discovery, practice), self-generated actions or performance (e.g., demonstrating competence, success or failure), other people (e.g., team members, patient), and outcomes (anticipated or known).

In terms of antecedents for emotions, analyses revealed that across both the real case and simulation, factors such as *task importance, performance/outcome, experience/skills*, and *difficulty level* were commonly reported by participants in association with emotions. In terms of differences, *team dynamics* and *time constraints* were unique to the real case, whereas *task* *novelty* and *fidelity* were unique to the simulation (although the antecedent of *task initiation* described in the real case may be a similar concept). The association between these antecedents and emotions was as expected, with high task difficulty associated with negative emotions (e.g., frustration), low task difficulty associated with positive emotions (e.g., relief), expectations for (or reactions to) successful performance associated with positive emotions (e.g., pride), and high level of experience/skills was associated with positive emotions (e.g., enjoyment). Conversely, expectations for (or reactions to) unsuccessful performance were associated with negative emotions (e.g., disappointment) and low levels of experience/skills were associated with negative emotions (e.g., confusion). High time constraints were associated with negative emotions (fear), whereas low time constraints were associated with positive emotions (e.g., relaxation). Positive team dynamics were associated with positive emotions (e.g., gratitude), whereas negative team dynamics were associated with negative emotions (e.g., anxiety). Fidelity and novelty antecedents were more mixed, although states of surprise and curiosity were linked to these antecedents. Task importance was also mixed: Most participants reported that the task was very important, but this was associated with negative emotions when success was uncertain and positive emotions when success was more certain.

Regarding regulation strategies, participants discussed strategies they recalled using in surgical settings to manage their emotional states. A range of emotion regulation strategies were described, including *reappraisal* (e.g., changing or reframing the way an event is perceived), *response modulation* (e.g., modulating speech or breathing, suppressing expression), *situation-modification* (e.g., problem-solving, seeking help) *competency-building* (increase knowledge or skills), and *attention-focused* strategies (e.g., increasing concentration, distracting oneself), *savoring* (e.g., appreciating a positive state to extend it), *mindfulness/acceptance* (e.g., acknowledging and recognizing emotion without evaluative judgment), and *regulation of others* (e.g., directing team members' attention or motivation). As to the purpose of emotion regulation, in most cases, participants maintained that they deployed these strategies with the purpose of decreasing the intensity or duration of negative emotional states (e.g., anxiety, confusion, frustration, disappointment). In some cases, participants discussed using regulations strategies to increase or maintain positive emotions (e.g., to keep focused, remain relaxed; learn more about case to increase curiosity and enjoyment). Less commonly, participants recalled using regulation strategies to intentionally increase negative emotions—or perhaps physiological arousal (e.g., increase stress to concentrate). These variations likely relate to a range in subjective perceptions regarding the "optimal" emotional state for successful performance. Although participants most commonly described ideal states in terms of positive emotions (excitement, enjoyment, calmness), several participants noted that some degree of stress helped them to focus their attention on the task at hand.

In addition to the use of various emotion regulation strategies, there was also a wide range in the degree to which participants were aware of their emotional states and valued using strategies. For instance, some participants maintained that they did not regulate their emotions (e.g., negative emotions may be adaptive, ideal states may occur naturally) or they considered their emotions to be irrelevant. For example, as one participant stated: "I don't [regulate them] actually, because I think if you care, and you're a junior, it's going to be there, the undertone of [anxiety]." Another participant noted: "Well I think if I have to do it [surgical procedure], I have to do it. Emotion has nothing to do with the task." However, it could be argued that these individuals were using the strategies of *distancing* or *acceptance*—perhaps in a more automated than deliberate manner. In comparison, some participants spoke at length about managing their emotions and its value. As one participant stated: "I'm actually really good at controlling my emotions. It's one of the things that I try to keep them in check; see what I can do to make the situation better." In contrast, other participants indicated that they lacked awareness of their emotions *or* insight into their use of specific regulatory strategies. As one participant summarized: "I'm not in tune with my emotions at all, so I just go and do what I have to do." Another participant remarked: "I'm not sure that I'm cognizant of how I do that [regulate emotions]. It really has to do with repetition." Few participants listed regulation strategies specific to the simulation task, and instead recalled using them for real cases in the operating room. This finding may be due, in part, to participants believing that their negative emotions were not sufficiently intense to warrant management. As one participant noted after the simulation: "There is no time to first develop the emotions." This finding may also be due to key differences between features of the two conditions that, according to some participants, included consequences, time pressure, emotional intensity, and evaluative pressures.

In terms of the fidelity of the vein model simulation, participants noted aspects that they considered to be the most realistic including: anatomy (positioning of the vein and tissues), tissue quality/feel, and procedural/technical steps. One of the most commonly mentioned suggestions for improvement was increasing the number and quality of branching veins (closer approximation of fluid) as separating these from the greater saphenous vein can be one of the most time-intensive aspects of the procedure in a real case, requiring careful technique to control bleeding and prevent damage). Other suggestions included improving tissue quality (increasing density and stability for suturing), reducing vein size, and more broadly, increasing complexity (suggested by more experienced participants). Regarding the utility of the simulation for training, most participants reported that it provided a useful opportunity for medical students and junior residents to improve basic surgical skills (suturing, knot-tying) and instrument handling, and to learn more about the vein harvest procedure and technical steps involved. Participants also noted that it can be useful for those with more experience to practice more complex techniques, improve efficiency, or prepare for more challenging aspects of the procedure (e.g., trouble-shooting when time is not a concern, handling fragile tissue). Finally, some participants noted that the simulation could help trainees build confidence and stay calm (i.e., emotional preparedness for the real case). In the following section, we elaborate on this concept of emotional preparedness, the implications for surgical training, and directions for future work. We start by discussing the main findings of this study.

Discussion

The purpose of this study was to examine the nature of emotions among surgical trainees and experts during an authentic technical procedure within the operating room to understand how emotions relate to several facets of learning and achievement including task appraisals (i.e., control value perceptions), task processes (i.e., regulation strategies), task performance (i.e., technical performance), and task outcomes (i.e., self-efficacy, satisfaction, perceived competence). Differences in these variables were also examined according to experience level (i.e., expert versus novice) and task phases (i.e., before, during, after procedure). These relations and mean differences were examined in both the real case and simulation condition to examine whether different patterns emerged. Accordingly, our research questions for both conditions asked: (1) Do fluctuations in emotional states emerge across multiple points in time for a surgical procedure? (2) Do differences in emotions emerge across levels of surgical experience? (3) How do emotional states relate to task appraisals, processes, performance, and outcomes? The discussion begins by addressing the three key questions and related hypotheses. For each of the findings described, similarities and differences between the two conditions are discussed.

Implications of these findings are then discussed, followed by the limitations and future directions of this work.

Validity of Training Conditions and Measures

Findings indicate that the simulation model designed for the present study represented a valid learning experience. Raters were able to accurately discriminate expert and novice performance using the simulation model. Participants reported that the model was realistic and useful for practicing their technique and improving skill. In fact, greater increases in proficiency reportedly occurred in the simulation compared to the real case. Perhaps the lower-stakes situation allowed novices to assume greater autonomy without concern over negative consequences. Given that emotions were less intense in this condition, this may have helped to reduce demands on working memory capacity (see Beilock, 2008; Mattarell-Micke et al., 2011) and allow trainees to focus on learning the step-by-step procedural components of the task rather than diverting their attention toward anticipated outcomes.

In terms of the measures used in this study, scale reliabilities were sufficiently high with the exception of the positive deactivating emotions (i.e., relief, relaxation) in the simulation, which were not related to one another. The subscale for negative deactivating emotions also required the removal of boredom in the simulation as it was not sufficiently correlated with the remaining discrete states within this category (hopelessness, disappointment, sadness). In both cases, it is likely that these discrete states are activated at different levels of experience. Experts may feel bored or relaxed because they can easily handle the task, whereas novices may feel disappointed or relieved due to concerns over performance. Both negative and positive *activating* emotions seemed to be the most consistently reliable compared to *deactivating* emotions.

Emotions across Time, Experience Level, and Condition

Results revealed that the most intensely experienced discrete positive emotions across both conditions included enjoyment/happiness, relaxation, pride, hopefulness, and relief. Compassion was more intense in the real case than the simulation, which is likely due to the real condition involving a patient as opposed to the use of a mannequin in the simulation. The most intensely experienced discrete negative emotions across both conditions included anxiety, fear, frustration, disappointment, and confusion. Consistent with our hypotheses, these negative emotions were more intense in the real case than the simulation, which was expected given the high-stakes nature of the real case. Results were largely similar for quantitative and qualitative analyses. These findings demonstrate that the most intensely activated emotions include a diverse range of activating and deactivating states that represent emotions from social, epistemic, achievement, and basic emotion taxonomies.

Results also revealed several changes that occurred in emotions over time and experience level. For instance, in both conditions, positive deactivating emotions (e.g., relief, relaxation) increased in intensity over time (highest after the task) and negative activating emotions decreased in intensity over time (lowest after the task) across all participants but were more intensely experienced by novices particularly during the task compared to experts. In the real case, negative deactivating emotions were more intense among novices compared to experts after the task. These findings, which are consistent with our hypothesis, demonstrate that novices are more susceptible to experience negative emotions, such as confusion and frustration during the task, as well as hopelessness and disappointment after the task, which has implications for instructional designs aimed to facilitate positive learning experiences. Collectively, these findings provide insight into the nature of emotions within surgical training and suggest that future research should examine a comprehensive range of emotional states, attending to possible change over time, task condition, and experience level. The following section addresses how these differences in emotional states arise and how they relate to performance and task outcomes.

Relations and Roles of Task Appraisals, Emotions, Performance, and Outcomes

In terms of appraisals, as predicted, control was positively related to positive emotions, negatively related to negative emotions, and positively related to performance and task outcomes within the real case. Overall, control was related to more variables than value and seemed to be a more powerful antecedent to emotion and subsequent outcomes, particularly in the real case. One possible reason is that the relationship between value, emotions, and outcomes is more complex; according to CVT, high levels of value can promote positive or negative emotions depending on whether or not the outcome is expected to be successful and the degree of certainty about this result. This was consistent with findings from qualitative analyses. The importance of the task was often associated with negative emotions when success was uncertain and positive emotions when success was more certain. In a similar vein, different facets of value (e.g., task importance versus intrinsic interest, Wigfield & Eccles, 2000) may have differential effects on emotions.

Another possible explanation is that value was relatively high among all participants and did not have the same degree of variability between experts and novices as the control appraisal (i.e., a ceiling effect). The categories of antecedents that emerged from qualitative analyses identified control and value as key antecedents. Namely, the categories of experience/skill level, task difficulty, team dynamics, and time constraints align closely with a control appraisal, whereas the categories of task importance and performance/outcomes align closely with the value appraisal. The antecedents novelty and fidelity, which emerged uniquely in the simulation condition, are not included in CVT, but align with the concept of novelty/unexpectedness that is included within the component process model of emotions (Scherer, 2009).[25] Although it may be argued that novelty and fidelity are related to how interesting or valuable a task is perceived to be, these findings suggest that simulation researchers should more closely examine the relevance of these antecedents, as the number of appraisals needed to fully account for emotional experiences may vary across contexts. Although some relationships were no longer statistically significant when controlling for prior experience, the relations between control, emotions, and task outcomes remained largely stable within the real case, suggesting that these relations exist beyond prior experience. The path analysis provided further support for CVT by demonstrating that the relationship between control and performance was mediated by emotion.

With respect to emotions relations to performance and task outcomes, results revealed that in both conditions, negative activating emotions had the strongest relationship with performance and task outcomes; the more individuals reported feeling emotions such as confusion, frustration, or anxiety, the more time it took them to complete the task, the lower their technical performance, the lower their satisfaction and post-task self-efficacy. This is consistent with previous research demonstrating anxiety is negatively related to performance in the health professions (e.g., Harvey et al., 2012, Hunziker et al., 2011; LeBlanc et al., 2005, 2012). Negative deactivating emotions were also negatively related to task outcomes in both conditions; in other words, the more individuals felt emotions such as boredom, disappointment, or hopelessness, the less proficiency gains they reported, the lower self-efficacy they felt after the task, and the less satisfied they were with their performance. On the other hand, positive deactivating emotions (relaxation and relief) were positively related to task outcomes in the real case. These findings suggest that positive emotions may be an important factor promoting more constructive views of the learning experience and future performance, whereas negative emotions may play a detrimental role in performance and lead to more pessimistic views about

[25] This model provides a framework for understanding emotions in everyday life unbounded by achievement context

the learning experience, which is consistent with our hypotheses. This is important from an instructional standpoint, as feelings of hopelessness and negativity about task outcomes may be consequential for motivation (e.g., seeking opportunities to increase competencies). As previously noted, these states are more intensely experienced by novices.

With respect to regulation strategies, results revealed that within the real case, the more individuals reported using cognitive reappraisal strategies, the higher their performance and the lower their physiological arousal (assessed through electrodermal activity). In contrast, suppression strategies were unrelated to performance. This is consistent with the literature in psychology and affective sciences, which has shown that changing the way an event is perceived is more adaptive than concealing emotions (Cutuli, 2014). Moreover, these findings indicate regulation is linked to physiological changes, which may bear on motor movements (precision, accuracy) and ultimately performance. These two strategies also emerged prominently in the qualitative analyses of interviews, although other regulation strategies were also described in participant responses. For instance, the broader category of *response modulation* includes suppressing an emotion, but also includes changing the physiological response (e.g., relaxation, breathing techniques), which may be more adaptive. Participants also described strategies such as problem-solving, increasing knowledge/skills, help-seeking, distraction, and concentration techniques, which largely map onto Gross' (2015) process model of emotion regulation, as well as regulation strategies described in CVT, including competence-oriented strategies, which Pekrun and Perry (2014) contend is particularly useful for achievement settings. Other regulation strategies described, such as *savoring, acceptance*, *regulation of others* did not fit as clearly within these frameworks, although several similar strategies have been described in the psychology literature (e.g., Mitmansgruber et al., 2009; Tugade & Fredrickson, 2007). This is consistent with previous research in surgical environments, which has found a range of coping strategies reportedly used to deal with stress (Arora et al., 2009; Hassan et al., 2006; Wetzel et al. 2010), several of which have been linked to performance.

Reappraisal and suppression strategies in particular have received the most attention in the emotion regulation literature, and although the findings from this research attest to their relevance, the qualitative results suggest that other regulation strategies should also be considered. Furthermore, the range in participants' awareness of emotions and degree of intentionality behind their regulation efforts suggest that further research should explore facets of regulation related to identification (recognizing need to change an emotional state), selection (choosing a strategy), and implementation (effectively deploying the strategy, see Gross, 2015). The reason why regulation did relate to performance in the simulation is likely due to the lack of intensity of negative emotions experienced in the simulation—and perhaps the lack of need to regulate. Given its importance in clinical practice, this suggests a need to include emotion regulation in surgical training. Although it has been argued that self-regulation plays a critical role in medical education (Butler & Brydges, 2013), the role of emotion regulation has received less attention than strategies directed towards cognitive facets of learning and performance.

Theoretical, Methodological, Practical Contributions

This study is the first of its kind to demonstrate the importance of emotions for surgical tasks by testing CVT. The findings from this research provide a better understanding of the nature and role of emotions by assessing emotion intensity over time, experience, and conditions within surgical training. Unlike previous work, this study examines the role of *state* emotions for *authentic* surgical tasks and their relations to performance. This represents a departure from previous work that has typically focused on trait-based emotions, rather than feelings experienced in the moment. Moreover, prior to this research, little data existed regarding the nature of emotions in the OR for a real case. In terms of theoretical contributions, this research provides support for control-value theory and its utility as a framework to understand and predict the role of emotions in medical education. Our hypotheses grounded in control-value theory were largely supported by the findings, particularly for the real case, and most consistently with respect to control appraisals. Based on the findings from this study, adding novelty as an appraisal in future empirical inquiry may provide a more comprehensive understanding of the onset of emotions within simulation environments.

In terms of methodological contributions, the use of multi-modal data channels (self-report, physiological data, observations) and mixed-methods analyses (quantitative and qualitative) allowed us to triangulate findings and identify novel patterns and areas for further investigation in future research. The broad range of emotions measured in this study also provides a unique contribution. Prior research has focused specifically on anxiety and coping, which emphasize a bias toward negative emotions. By including emotions that represent a range of taxonomies, we were able to measure a more comprehensive array of emotional experiences. Conducting qualitative analyses of interview responses allowed us to gain a richer understanding—from the participants' perspective—of antecedents and regulation strategies, aside from the more popularized constructs of reappraisal and suppression.

In terms of practical implications for surgical training, the simulation vein harvest model we designed for this study provides a feasible and useful model for training residents to practice the vein harvest procedure and assess performance. Prior to this study, the vein harvest technique had received very little attention in the surgical education literature. Instead, attention has typically been devoted to more complex procedures. However, given the importance of this procedure for effective outcomes in cardiac surgery, and its role in training, the findings from this research can be used to inform the design of simulation training and interventions that aim to promote more adaptive emotions for vein harvest training, as well as other surgical procedures and training environments. We elaborate on these implications next.

Educational Implications

In terms of simulation training, one direct application is to integrate the simulation model into the cardiac surgery-training curriculum (see Baker, 2012) to prepare trainees to conduct the saphenous vein harvest procedure in the operating room. Another possibility is to use this type of model to assess competencies before conducting this procedure during a real case. The results from qualitative analyses identify areas to further develop the model, such as increasing the complexity of the procedure for more experienced participants and simulating team dynamics to more closely approximate interactions in the OR.

More broadly, the findings from this research suggest that simulation designers should not only consider the *technical fidelity* (i.e., extent to which materials and techniques approximate real tissues or procedures) but also the *psychological fidelity* (Kozlowski & DeShon, 2004), which refers to the extent to which simulations elicit psychological processes important for performance in the real-world, including emotion and regulation. This is consistent with Hamstra et al.'s (2014) position that fidelity is multifaceted and should be considered not only in terms of *physical* resemblance but also *functional* task alignment to real-world practice. Although the major findings (e.g., importance of control, relationship between negative emotions, performance, and task outcomes) were consistent across conditions, discrepancies were likely due to the lack of intensity of emotional experience in the simulation. From a deliberate practice perspective (Reznick & Macrae, 2006), less intense emotions can be useful at early stages of procedural skill development to focus on the technique without significant demands to cognitive load or concerns about negative consequences. On the other hand, considering that high-stress situations are likely to occur in the real case, trainees and experts must eventually learn to manage their emotions. Increasing the degree of simulation fidelity gradually over the learning trajectory may help to optimize learning efficiency and clinical transfer (Brydges et al., 2010).

Given the potential consequences of negative activating emotions in particular, efforts should be made to monitor their intensity, especially among novices. Although positive emotions alone may not account for higher performance, positive emotions may help to promote learning gains and constructive views of the learning experience. In some cases, this may involve increasing the intensity and duration of positive emotions. In other instances, this may involve decreasing the intensity and duration of negative emotions to attain a more adaptive state.

To promote more adaptive emotional states for simulations and real cases in the operating room, one approach is to target efforts from a time sequence perspective in line with Gross' (2015) Process Model of Emotion Regulation. For example, a focus on setting clear goals, providing instructions, and outlining mechanisms for support before the task can enhance perceived control that, as this research suggests, can help to promote more adaptive emotional states and enhance performance. During the task, team support and co-regulation (see Järvenoja & Järvelä, 2009; Järvenoja, Volet, & Järvelä, 2013) may help to manage states of confusion, frustration, or anxiety if complications arise. After the task, providing specific and immediate feedback, as well as further opportunities to increase competence (e.g., simulation training), could improve perceptions of control prior to conducting the real procedure. These considerations are consistent with control-value theory, which postulates that subjective appraisals of control and value events are influenced by a host of features embedded in the social environment such as the quality of instruction, achievement expectations, and cultural values (Pekrun et al., 2006). At the same time, trainees can learn to regulate their own emotions by monitoring their emotions and deploying effective strategies (Gross, 2015). To facilitate this process, educational interventions aimed at understanding the role of emotion and regulation may be beneficial for surgical trainees and professionals. As Brydges et al (2015) suggest, effective regulation may require simulation training and instructional designs that help trainees to effectively regulate, rather than placing the responsibility solely on the individual.

Limitations and Future Directions

There are several limitations to the present study that could be addressed in future research. To begin, the study involved a relatively small sample size. Although our sample was highly specialized and reflective of the number of novices and experts in this medical speciality, a multi-site study would allow us to determine whether the results are robust across a larger group of participants. Similarly, sampling a broader range of surgical procedures at higher levels of complexity would allow us to test the generalizability of these results to other tasks and training conditions. In addition, although we assessed change over time for each task, future research could assess change on a larger scale as participants engage in repeated practice of surgical procedures and transfer tasks.

Secondly, although we assessed task outcomes through self-reported responses, the use of pre-task, post-task, and retention tests of performance, knowledge, and perceptions would allow us to tease apart the effect of training conditions on learning gains and emotions, as well as provide a more sensitive measure of experience level. In a similar vein, using an experimental design would allow us to more directly test the impact of training conditions on emotions, whether simulation leads to improvements in clinical practice through enhanced emotion preparedness and effective regulation, and at what point in training does enhanced psychological fidelity optimize learning. Certainly the challenges of implementing these research designs in a real-world training environment should not be overlooked (see Sullivan, 2011). However, the findings from this study provide the groundwork for implementing educational interventions designed to promote more adaptive emotional states for both real case and simulation training.

Technical performance could also be examined at a finer level of granularity (e.g., careful handing of tissue versus efficiency) as there are other aspects of surgical performance that can impact patient outcomes and may be affected by emotions, such as decision-making and team communication. Considering the team-based nature of surgery, examining the role of emotions from a social perspective (e.g., emotion contagion, shared regulation) may help to provide a richer understanding of how the socio-cultural environment can impact emotions during surgery. In addition, although we focused on grouping emotions according to control-value theory (i.e., valence and arousal), future research should also examine the orthoganality of discrete emotion states, the types of emotions that commonly co-occur (e.g., moderate anxiety combined with high curiosity), and the factor structure of experienced emotions, particularly in light of positive and negative deactivating emotions demonstrating lower internal reliability in this study due to divergent composite items (e.g., relaxation vs. relief).

In addition, a closer examination of task processes may help us to better understand the mechanisms underlying the link between emotion and performance. It may be the case, for instance, that emotions impact technical performance by affecting motor movements (see Coombes & Cauraugh & Christopher, 2007; Naugle et al., 2012) via changes in physiological arousal. It may also be the case that emotions impact technical performance by affecting attentional processes (e.g., over or under-attending to critical areas of interest) or cognitive judgments (Blanchette & Richards, 2010; Starcke & Brand, 2012) due to strains on working memory capacity (i.e., cognitive overload, see Young et al., 2014). For novices, this may interfere with execution of step-by-step procedures; for experts, this may involve regressing from an automatized to procedural approach (see Yu, 2015).

Finer-grained analysis of eye-tracking data would help to test these hypotheses. In the present study, eye-tracking recordings were analyzed exclusively as a channel for performance and time on task; however, gaze behavior data captured in these recordings can be mined to identify more detailed behavioral profiles for experts and novices in relation to emotions (see Alsop & Gray, 2014; Bebko et al., 2011). In a similar vein, data from electrodermal activity can be mined to identify physiological traces of emotion regulation over time (see Kreibig, 2010). By pursuing these types of analyses and avenues for further inquiry, our goal is to advance our understanding of the role of emotions in surgical training. Reznick and MacRae (2006) note that volume of surgeries alone does not fully account for variations in performance among surgeons; the findings from the current study offer compelling evidence for the role of emotion.

References

Affectiva. (2014). Q U Manual. Available at: http://qsensor-support.affectiva.com/customer/portal/articles/1539375-q-user-manual. Accessed November 2015.

Aggarwal, R., Grantcharov, T., Moorthy, K., Milland, T., & Darzi, A. (2008). Toward feasible, valid, and reliable video-based assessments of technical surgical skills in the operating room. *Annals of Surgery, 247,* 372–379.

Aggarwal, R., Mytton, O. T., Derbrew, M., Hananel, D., Heydenburg, M., Issenberg, B., … Reznick, R. (2010). Training and simulation for patient safety. *Quality and Safety in Health Care*, *19*(Suppl 2), i34–i43.

Agha, R. A., Fowler, A. J., & Sevdalis, N. (2015). The role of non-technical skills in surgery. *Annals of Medicine and Surgery*, *4*(4), 422–427.

Allsop, J., & Gray, R. (2014). Flying under pressure: Effects of anxiety on attention and gaze behavior in aviation. *Journal of Applied Research in Memory and Cognition*, *3*(2), 63–71.

American College of Cardiology and American Heart Association Task Force. (2004). Guideline update for coronary artery bypass graft surgery. *Journal of the American College of Cardiology, 44*, 1146–1154.

Angie, A. D., Connelly, S., Waples, E. P., & Kligyte, V. (2011). The influence of discrete emotions on judgement and decision-making: A meta-analytic review. *Cognition and Emotion*, *25*(8), 1393–1422.

Antiel, R. M., Van Arendonk, K. J., Reed, D. A., Terhune, K. P., Tarpley, J. L., Porterfield, J. R., … Farley, D. R. (2012). Surgical training, duty-hour restrictions, and implications for meeting the Accreditation Council for Graduate Medical Education core competencies: views of surgical interns compared with program directors. *Archives of Surgery*, *147*(6), 536–541.

Arora, S., Sevdalis, N., Suliman, I., Athanasiou, T., Kneebone, R., & Darzi, A. (2009). What makes a competent surgeon? Experts' and trainees' perceptions of the roles of a surgeon. *The American Journal of Surgery*, *198*(5), 726–732.

Arora, S., Nestel, D., Sevdalis, N., Woloshynowych, M., Darzi, A., & Kneebone, R. (2010a). The impact of stress on surgical performance: A systematic review of the literature.

Surgery, *147*(3), 318–330.

Arora, S., Sevdalis, N., Aggarwal, R., Sirimanna, P., Darzi, A., & Kneebone, R. (2010b). Stress impairs psychomotor performance in novice laparoscopic surgeons. *Surgical Endoscopy and Other Interventional Techniques, 24,* 2588–2593.

Aucar, J. A., Groch, N. R., Troxel, S. A., & Eubanks, S. W. (2005). A review of surgical simulation with attention to validation methodology. *Surgical Laparoscopy, Endoscopy & Percutaneous Techniques*, *15*(2), 82–89.

Babineau, T. J., Becker, J., Gibbons, G., Sentovich, S., Hess, D., Robertson, S., & Stone, M. (2004). The "cost" of operative training for surgical residents. *Archives of Surgery*, *139*(4), 366–370.

Baker, G. R., Norton, P. G., Flintoft, V., Blais, R., Brown, A., Cox, J., … Tamblyn, R. (2004). The Canadian Adverse Events Study: the incidence of adverse events among hospital patients in Canada. *Canadian Medical Association Journal*, *170*(11), 1678–1686.

Baker, C. J., Sinha, R., & Sullivan, M. E. (2012). Development of a cardiac surgery simulation curriculum: From needs assessment results to practical implementation. *The Journal of Thoracic and Cardiovascular Surgery*, *144*(1), 7–16.

Bandura, A. (2006). Guide for construction self-efficacy scales. In F. Pajares & T. Urdan (Eds.). *Self-Efficacy Beliefs of Adolescents,* (pp. 307–337). Greenwich, CT: Information Age Publishing.

Bandyopadhyay, D., Pammi, V. C., & Srinivasan, N. (2013). Role of affect in decision making. *Progress in brain research*, *202*, 37–53.

Bebko, G. M., Franconeri, S. L., Ochsner, K. N., & Chiao, J. Y. (2011). Look before you regulate: Differential perceptual strategies underlying expressive suppression and cognitive reappraisal. *Emotion*, *11*(4), 732–742.

Beilock, S. L. (2008). Math performance in stressful situations. *Current Directions in Psychological Science, 17,* 339–343.

Berguer, R., Smith, W. D., & Chung, Y. H. (2001). Performing laparoscopic surgery is significantly more stressful for the surgeon than open surgery. *Surgical Endoscopy, 15, 1204*–1207.

Blanchette, I., & Richards, A. (2010). The influence of affect on higher level cognition: A review of research on interpretation, judgement, decision making and reasoning. *Cognition and*

Emotion, *24*(4), 561–595.

Blum, M. G., Powers, T. W., & Sundaresan, S. (2004). Bronchoscopy simulator effectively prepares junior residents to competently perform basic clinical bronchoscopy. *The Annals of Thoracic Surgery*, *78*(1), 287–291.

Bridges, M., & Diamond, D. L. (1999). The financial impact of teaching surgical residents in the operating room. *The American Journal of Surgery*, *177*(1), 28–32.

Brydges, R., Carnahan, H., Rose, D., Rose L., & Dubrowski, A. (2010). Coordinating progressive levels of simulation fidelity to maximize educational benefit. *Academic Medicine, 85*(5), 806–812.

Brydges, R., Manzone, J., Shanks, D., Halata, R., Hamstra, S. J., Zendejas, B., & Cook, D. A. (2015). Self-regulated learning in simulation-based training: A systematic review and meta-analysis. *Medical Education, 49*, 368–378.

Buckley, C. E., Kavanagh, D. O., Traynor, O., & Neary, P. C. (2014). Is the skillset obtained in surgical simulation transferable to the operating theatre? *The American Journal of Surgery*, *207*(1), 146–157.

Butler, D. L., & Brydges, R. (2013). Learning in the health professions: what does self-regulation have to do with it? *Medical Education, 47,* 1057–1059.

Chi, M. T. H. (2006). Two approaches to the study of experts' characteristics. In K. A. Ericsson, N. Charness, P. Feltovich, & R. Hoffman (Eds.), *The Cambridge handbook of expertise and expert performance* (pp. 12130), New York, NY: Cambridge University Press.

Coombs, S. A., Cauraugh, J. H., & Jannell, C. M. (2007). Emotional State and Initiating Cue Alter Central and Peripheral Motor Processes. *Emotion, 7*(2), 275–284.

Creedon, P. J., & Hayes, A. F. (2015, May). *Small sample mediation analysis: How far can you push the bootstrap?* Paper presented at the annual conference of the Association for Psychological Science, New York, NY.

Cutuli, D. (2014). Cognitive reappraisal and expressive suppression strategies role in the emotion regulation: an overview on their modulatory effects and neural correlates. *Frontiers in Systems Neuroscience, 8,* 1–6.

de Montbrun, S. L., & MacRae, H. (2012). Simulation in Surgical Education. *Clinics in Colon and Rectal Surgery*, *25*(3), 156–165.

de Vries, E. N., Ramrattan, M. A., Smorenburg, S. M., Gouma, D. J., & Boermeester, M. A. (2008). The incidence and nature of in-hospital adverse events: a systematic review. *Quality & Safety in Health Care*, *17*(3), 216–223.

de Vries, E. N., Prins, H. A., Crolla, R. M. P. H., den Outer, A. J., van Andel, G., van Helden, S. H., … Boermeester, M. A. (2010). Effect of a comprehensive surgical safety system on patient outcomes. *The New England Journal of Medicine*, *363*(20), 1928–1937.

Dimick, J. B., Chen, S. L., Taheri, P. A., Henderson, W. G., Khuri, S. F., & Campbell, D. A. (2004). Hospital costs associated with surgical complications: A report from the private-sector National Surgical Quality Improvement Program. *Journal of the American College of Surgeons*, *199*(4), 531–537.

Doty, D. B., & Doty, J. R. (2012). Coronary artery bypass graft. In D. B Doty & J. R. Doty (Eds.), *Cardiac surgery: Operative technique (2nd Ed.)* (pp. 393–431) Philadelphia, PA: Elsivier.

Dunnington, G. L. (1996). The art of mentoring. *American Journal of Surgery*, *171*(6), 604–607.

Eccles, J. S., Adler, T. F., Futterman, R., Goff, S. B., Kaczala, C. M., Meece, J. L., & Midgley, C. (1983). Expectancies, values, and academic behaviors. In J. T. Spence (Ed.), *Achievement and achievement motivation* (pp. 75–146). San Francisco, CA: W. H. Freeman.

Farkouh, M. E., Domanski, M., Sleeper, L. A., Siami, R. S., Dangas, G., Mack, M., … Fuster, V. (2012). Strategies for multivessel revascularization in patients with diabetes. *New England Journal of Medicine, 367*, 2375–2384.

Fawver, B., Amano, S., Hass, C. J., & Janelle, C. M. (2012). Active control of approach-oriented posture is influenced by emotional reactions. *Emotion*, *12*(6), 1350–1361.

Flin, R., Yule, S., Paterson-Brown, S., Maran, N., Rowley, D., & Youngson, G. (2007). Teaching surgeons about non-technical skills. *The Surgeon*, *5*(2), 86–89.

Fritz, M. S., & MacKinnon, D. P. (2007). Required sample size to detect the mediated effect. *Psychological Science, 18*, 233–239.

Gaba, D. M. (2004). The future vision of simulation in health care. *Quality and Safety in Health Care*, *13*(suppl 1), i2–i10.

Gawande, A. A., Zinner, M. J., Studdert, D. M., & Brennan, T. A. (2003). Analysis of errors reported by surgeons at three teaching hospitals. *Surgery*, *133*(6), 614–621.

Gélat, T., Coudrat, L., & Le Pellec, A. (2011). Gait initiation is affected during emotional conflict. *Neuroscience Letters*, *497*(1), 64–67.

Global Burden of Disease 2013 Mortality and Causes of Death, Collaborators (2015). Global, regional, and national age-sex specific all-cause and cause-specific mortality for 240 causes of death, 1990-2013: A systematic analysis for the Global Burden of Disease Study 2013. *The Lancet*, *385*, 117–171.

Greene, J. A., Yu, S. B., & Copeland, D. Z. (2014). Measuring critical components of digital literacy and their relationships with learning. *Computers & Education, 76,* 55–69.

Greene, J. A., Dellinger, K., Binbasaran Tuysuzoglu, B., & Costa, L. (2013). A two-tiered approach to analyzing self-regulated learning process data to inform the design of hypermedia learning environments. In R. Azevedo, & V. Aleven (Eds.), *International handbook of metacognition and learning technologies* (pp. 117–128). New York: Springer.

Greenberg, C. C., Regenbogen, S. E., Studdert, D. M., Lipsitz, S. R., Rogers, S. O., Zinner, M. J., & Gawande, A. A. (2007). Patterns of Communication Breakdowns Resulting in Injury to Surgical Patients. *Journal of the American College of Surgeons*, *204*(4), 533–540.

Gross, J. J. (2015). The extended process model of emotion regulation: Elaborations, applications, and future directions. *Psychological Inquiry, 26,* 130-137.

Gross, J. J., & John, O. P. (2003). Individual differences in two emotion regulation processes: Implications for affect, relationships, and well-being. *Journal of Personality and Social Psychology, 85,* 348-362.

Gulati, R., Rihal, C. S., & Gersh, B. J. (2009). The SYNTAX Trial A perspective. *Circulation: Cardiovascular Interventions*, *2*(5), 463–467.

Hamdorf, J. M., & Hall, J. C. (2000). Acquiring surgical skills. *The British Journal of Surgery*, *87*(1), 28–37.

Hamstra, S. J., Byrdges, R., Hatala, R., Zendejas, B., & Cook, D. A. (2014). Reconsidering fidelity in simulation-based training. *Academic Medicine, 89,* 387–392.

Hassan, I., Weyers, P., Maschuw, K., Dick, B., Gerdes, B., Rothmund, M., & Zielke, A. (2006). Negative stress-coping strategies among novices in surgery correlate with poor virtual laparoscopic performance. *British Journal of Surgery, 93,* 1554–1559.

Hayes, A. F., & Preacher, K. J. (2014). Statistical mediation analysis with a multicategorical independent variable. *British Journal of Mathematical and Statistical Psychology, 67,* 451–470.

Haynes, A. B., Weiser, T. G., Berry, W. R., Lipsitz, S. R., Breizat, A. H. S., Dellinger, E. P., … Gawande, A. A. (2009). A surgical safety checklist to reduce morbidity and mortality in a global population. *New England Journal of Medicine*, *360*(5), 491–499.

Hess, C. N., Lopes, R. D., Gibson, C. M., Hager, R., Wojdyla, D. M., Englum, B. R., … Alexander, J. H. (2014). Saphenous vein graft failure after coronary artery bypass surgery: Insights from PREVENT IV. *Circulation, 130,* 1445–1451.

Hull, L., Arora, S., Aggarwal, R., Darzi, A., Vincent, C., & Sevdalis, N. (2012). The impact of nontechnical skills on technical performance in surgery: A systematic review. *Journal of the American College of Surgeons*, *214*(2), 214–230.

Issenberg, S., McGaghie, W. C., Hart, I. R., Mayer, J. W., Felner, J. M., Petrusa, E. R., … Ewy, G. A. (1999). Simulation technology for health care professional skills training and assessment. *Journal of the American Medical Association*, *282*(9), 861–866.

Järvenoja, H., & Järvelä, S. (2009). Emotion control in collaborative learning situations: Do students regulate emotions evoked by social challenges. *British Journal of Educational Psychology, 79*(3), 463–481.

Järvenoja, H., Volet, S., & Järvelä, S. (2013). Regulation of emotions in socially challenging learning situations: An instrument to measure the adaptive and social nature of the regulation process. *Educational Psychology, 33*(1), 31–58.

Kairys, J. C., McGuire, K., Crawford, A. G., & Yeo, C. J. (2008). Cumulative operative experience is decreasing during general surgery residency: A worrisome trend for surgical trainees? *Journal of the American College of Surgeons*, *206*(5), 804–811.

Kenton, K. (2006). How to Teach and Evaluate Learners in the Operating Room. *Obstetrics and Gynecology Clinics*, *33*(2), 325–332.

Kerr, B., & O'Leary, J. P. (1999). The training of the surgeon: Dr. Halsted's greatest legacy. *The American Surgeon*, *65*(11), 1101–1102.

Kneebone, R. (2005). Evaluating clinical simulations for learning procedural skills: A theory-based approach. *Academic Medicine: Journal of the Association of American Medical Colleges*, *80*(6), 549–553.

Kotsis, S. V., & Chung, K. C. (2013). Application of the "see one, do one, teach one" concept in surgical training. *Plastic and Reconstructive Surgery*, *131*(5), 1194–1201.

Kozlowski, S. W. J. & DeShon, R. P. (2004). A psychological fidelity approach to simulation-based training: Theory, research, and principles. In E. Salas, L. R. Elliott, S. G. Schflett, & M. D. Coovert (Eds.), *Scaled worlds: Development, validation, and applications* (pp. 75–99). Burlington, VT: Ashgate Publishing.

Kreibig, S. (2010). Autonomic nervous system activity in emotion: A review. *Biological Psychology, 84,* 394-421.

Larsen, C. R., Oestergaard, J., Ottesen, B. S., & Soerensen, J. L. (2012). The efficacy of virtual reality simulation training in laparoscopy: A systematic review of randomized trials. *Acta Obstetricia et Gynecologica Scandinavica*, *91*(9), 1015–1028.

Lindeman, B. M., Sacks, B. C., Hirose, K., & Lipsett, P. A. (2014). Duty hours and perceived competence in surgery: Are interns ready? *Journal of Surgical Research*, *190*(1), 16–21.

Lindström, B. R., & Bohlin, G. (2011). Emotion processing facilitates working memory performance. *Cognition and Emotion*, *25*(7), 1196–1204.

Lingard, L., Espin, S., Whyte, S., Regehr, G., Baker, G. R., Reznick, R., … Grober, E. (2004). Communication failures in the operating room: an observational classification of recurrent types and effects. *Quality & Safety in Health Care*, *13*(5), 330–334.

MacRae, H., Regehr, G., Leadbetter, W., & Reznick, R. K. (2000). A comprehensive examination for senior surgical residents. *The American Journal of Surgery, 179*, 190–193.

Martin, J. A., Regehr, G., Reznick, R., Macrae, H., Murnaghan, J., Hutchison, C. & Brown, M. (1997). Objective structured assessment of technical skill (OSATS) for surgical residents. *British Journal of Surgery, 84*, 273–278.

Mattarella-Micke, A., Mateo, J., Kozak, M. N., Foster, K., & Beilock, S. L. (2011). Choke or thrive? The relation between salivary cortisol and math performance depends on individual differences in working memory and math-anxiety. *Emotion, 11,* 1000–1005.

Mitmansgruber, H., Beck, T. N., Höfer, S., & Schüber, G. (2009). When you don't like what you feel: Experiential avoidance, mindfulness and meta-emotion in emotion regulation. *Personality and Individual Differences, 46*, 448–453.

Mohr, F. W., Morice, M. C., Kappetein, A P., Feldman, T. E., Ståhle, E., Colombo, A., … Serruys, P. W. (2013). Coronary artery bypass graft surgery versus percutaneous coronary intervention in patients with three-vessel disease and left main coronary disease: 5-year follow-up of the randomised, clinical SYNTAX trial. *The Lancet, 381,* 629-638.

Nasca, T. J., Day, S. H., Amis, E. S., & ACGME Duty Hour Task Force. (2010). The new recommendations on duty hours from the ACGME Task Force. *The New England Journal of Medicine*, *363*(2), e3.

National Heart, Lung, and Blood Institute. (2012). What is coronary artery bypass grafting? Available at: http://www.nhlbi.nih.gov/health/health-topics/topics/cabg/. Accessed January, 2015.

Naugle, K. M., Joyner, J., Hass, C. J., & Janelle, C. M. (2010). Emotional influences on locomotor behavior. *Journal of Biomechanics*, *43*(16), 3099–3103.

Naugle, K. M., Coombes, S. A., Cauraugh, J. H., & Janelle, C. M. (2012). Influence of Emotion on the control of low-level force production. *Research Quarterly for Exercise and Sport*, *83*(2), 353–358.

O'Connor, T., Papanikolaou, V., & Keogh, I. (2010). Safe surgery, the human factors approach. *The Surgeon*, *8*(2), 93–95.

Palter, V. N., & Grantcharov, T. P. (2014). Individualized deliberate practice on a virtual reality simulator improves technical performance of surgical novices in the operating room: a randomized controlled trial. *Annals of Surgery*, *259*(3), 443–448.

Pekrun, R. (2006). The control-value theory of achievement emotions: Assumptions, corollaries, and implications for educational research and practice. *Educational Psychology Review, 18*, 315–341.

Pekrun, R., Frenzel, A. C., Goetz, T., & Perry, R. P. (2007). The control-value theory of achievement emotions: An integrative approach to emotions in education. In P. A. Schutz & R. Pekrun (Eds.), *Emotion in education* (pp. 13–36). San Diego, CA: Academic Press.

Pekrun, R., Goetz, T., Titz, W., & Perry, R. P. (2002). Academic emotions in students' self-regulated learning and achievement: A program of qualitative and quantitative research. *Educational Psychologist*, *37*, 91–105.

Pekrun, R., & Linnenbrink-Garcia, L. (2014). Introduction to emotions in education. In R. Pekrun & L. Linnenbrink-Garcia (Eds.), *International handbook of emotions in education* (pp. 1–10). New York: Francis & Taylor/Routledge.

Pekrun, R., Perry, R. P. (2014). Control-value theory of achievement emotions. In R. Pekrun & L. Linnenbrink-Garcia (Eds.), *International handbook of emotions in education* (pp. 120–141). New York: Francis & Taylor/Routledge.

Preacher, K. J., & Hayes, A. F. (2008). Contemporary approaches to assessing mediation in communication research. In A. F. Hayes, M. D. Slater, & L. B. Snyder (Eds.), *The Sage sourcebook of advanced data analysis methods for communication research* (pp. 13–54). Thousand Oaks, CA: Sage.

Quartz, S. R. (2009). Reason, emotion and decision-making: risk and reward computation with feeling. *Trends in Cognitive Sciences*, *13*(5), 209–215.

Reznick, R. K., & MacRae, H. (2006). Teaching surgical skills—Changes in the wind. *The New England Journal of Medicine, 355,* 2664-2669.

Reznick, R., Regehr, G., MacRae, H., Martin, J., & McCulloch, W. (1997). Testing technical skill via an innovative "bench station" examination. *American Journal of Surgery. 173*(3), 226–230.

Rogers Jr, S. O., Gawande, A. A., Kwaan, M., Puopolo, A. L., Yoon, C., Brennan, T. A., & Studdert, D. M. (2006). Analysis of surgical errors in closed malpractice claims at 4 liability insurers. *Surgery*, *140*(1), 25–33.

Sachdeva, A. K. (2002). Acquisition and maintenance of surgical competence. *Seminars in Vascular Surgery*, *15*(3), 182–190.

Sachdeva, A. K., Buyske, J., Dunnington, G. L., Sanfey, H. A., Mellinger, J. D., Scott, D. J., … Burns, K. J. (2011). A new paradigm for surgical procedural training for surgical procedural training: *Current Problems in Surgery*, *48*(12), 854–968.

Sedlack, R. E., & Kolars, J. C. (2004). Computer simulator training enhances the competency of gastroenterology fellows at colonoscopy: Results of a pilot study. *The American Journal of Gastroenterology*, *99*(1), 33–37.

Seymour, N. E., Gallagher, A. G., Roman, S. A., O’Brien, M. K., Bansal, V. K., Andersen, D. K., & Satava, R. M. (2002). Virtual reality training improves operating room performance. *Annals of Surgery*, *236*(4), 458–464.

Sharma, B., Mishra, A., Aggarwal, R., & Grantcharov, T. P. (2011). Non-technical skills assessment in surgery. *Surgical Oncology*, *20*(3), 169–177.

Shouhed, D., Gewertz, B., Wiegmann, D., & Catchpole, K. (2012). Integrating human factors research and surgery: A review. *Archives of Surgery*, *147*(12), 1141–1146.

Stahel, P. F., Sabel, A.L., Victoroff, M. S., Varnell, J., Boygle, D. J., Clarke, T. J., … Mehler, P. S. (2010). Wrong-site and wrong-patient procedures in the universal protocol era: Analysis of a prospective database of physician self-reported occurrences. *Archives of Surgery*, *145*(10), 978–984.

Starcke, K., & Brand, M. (2012). Decision making under stress: A selective review. *Neuroscience & Biobehavioral Reviews*, *36*(4), 1228–1248.

Stins, J. F., & Beek, P. J. (2011). Organization of voluntary stepping in response to emotion-inducing pictures. *Gait & Posture*, *34*(2), 164–168.

Stuart, E. A. (2010). Matching methods for causal inference: A review and a look forward. *Statistical Science, 25*(1), 1–21.

Sturm, L. P., Windsor, J. A., Cosman, P. H., Cregan, P., Hewett, P. J., & Maddern, G. J. (2008). A systematic review of skills transfer after surgical simulation training. *Annals of Surgery*, *248*(2), 166–179.

Sullivan, G. M. (2011). Getting off the "gold standard": Randomized controlled trials and education research. *Journal of Graduate Medical Education, 3(3),* 285-289.

Tendulkar, A. P., Victorino, G. P., Chong, T. J., Bullard, M. K., Liu, T. H., & Harken, A. H. (2005). Quantification of surgical resident stress "on call". *Journal of American College of Surgeons, 201*, 560–564

Tsuda, S., Scott, D., Doyle, J., & Jones, D. B. (2009). Surgical skills training and simulation. *Current Problems in Surgery*, *46*(4), 271–370.

Tugade, M. M., & Fredrickson, B. L. (2007). Regulation of positive emotions: emotion regulation strategies that promote resilience. *Journal of Happiness Studies, 8,* 311–333.

Undre, S., Koutantji, M., Sevdalis, N., Gautama, S., Selvapatt, N., Williams, S., … Vincent, C. (2007). Multidisciplinary crisis simulations: the way forward for training surgical teams. *World Journal of Surgery, 31,* 1843–1853.

Vincent, C., Neale, G., & Woloshynowych, M. (2001). Adverse events in British hospitals: preliminary retrospective record review. *British Medical Journal*, *322*(7285), 517–519.

Walter, A. J. (2006). Surgical education for the twenty-first century: beyond the apprentice model. *Obstetrics and Gynecology Clinics of North America*, *33*(2), 233–236.

Wetzel C. M., Kneebone R. L., Woloshynowych M., Nestel D., Moorthy K., Kidd J., & Darzi, A. (2006). The effects of stress on surgical performance. *American Journal of Surgery, 191,* 5–10.

World Health Organization. (2014). The top 10 causes of death. World health statistics 2012. Available at: http://www.who.int/mediacentre/factsheets/fs310/en/. Accessed October 2015.

Wigfield, A., & Eccles, J. S. (2000). Expectancy-value theory of achievement motivation. *Contemporary Educational Psychology 25,* 68–81.

Williams, R. G., & Dunnington, G. L. (2004). Accreditation Council for Graduate Medical Education core competencies initiative: the road to implementation in the surgical specialties. *The Surgical Clinics of North America*, *84*(6), 1621–1646.

Wolf, O. T. (2009). Stress and memory in humans: Twelve years of progress? *Brain Research*, *1293*, 142–154.

Yamamoto, A., Hara, T., Kikuchi, K., & Fujiwara, T. (1999). Intraoperative stress experienced by surgeons and assistants. *Ophthalmic Surgery and Lasers, 30*, 27–30.

Yiend, J. (2010). The effects of emotion on attention: A review of attentional processing of emotional information. *Cognition and Emotion*, *24*(1), 3–47.

Young, J. Q., Van Merrienboer, J., Durning, S., & ten Cate, O. (2014). Cognitive load theory: Implications for medical education: AMEE Guide No. 86, *Medical Teacher, 36*, 371-384.

Yu, R. (2015). Choking under pressure: The neuropsychological mechanism of incentive-induced performance decrements. *Frontiers in Behavioral Neuroscience, 9,* 1–8.

Yule, S., Flin, R., Paterson-Brown, S., & Maran, N. (2006a). Non-technical skills for surgeons in the operating room: A review of the literature. *Surgery*, *139*(2), 140–149.

Yule, S., Flin, R., Paterson-Brown, S., Maran, N., & Rowley, D. (2006b). Development of a rating system for surgeons' non-technical skills. *Medical Education*, *40*(11), 1098–1104.

Zenati, M. A., Shroyer, A. L., Collins, J. F., Hattler, B., Ota, T., Almassi, G. H., … Sonel, A. F. (2011). Impact of endoscopic versus open saphenous vein harvest technique on late coronary artery bypass grafting patient outcomes in the ROOBY (Randomized On/Off Bypass) Trial. *The Journal of Thoracic and Cardiovascular Surgery*, *141*(2), 338–344.

Table 4.
Accuracy between Performance Score and Experience Level (Real Case)

Performance score (rating scale)	Experience level (expert/novice)	Correct classification (mid-way point)	Correct classification (median-split)
4.42	Expert	Consistent	Consistent
2.75	Novice	Inconsistent	Consistent
3.50	Expert	Consistent	Consistent
3.42	Expert	Consistent	Consistent
3.50	Expert	Consistent	Consistent
2.83	Expert	Consistent	Inconsistent
1.00	Novice	Consistent	Consistent
3.17	Expert	Consistent	Consistent
1.08	Novice	Consistent	Consistent
1.50	Novice	Consistent	Consistent
1.08	Novice	Consistent	Consistent
1.25	Novice	Consistent	Consistent
3.17	Expert	Consistent	Consistent
2.58	Novice	Inconsistent	Consistent
4.00	Expert	Consistent	Consistent
3.25	Expert	Consistent	Consistent
4.33	Expert	Consistent	Consistent
4.58	Expert	Consistent	Consistent
2.75	Novice	Inconsistent	Consistent

Note. For mid-way point classification, performance scores above 2.5 classified as *high*; scores below as *low*. For median-split classification, performance scores above median score (3.167) classified as *high*; scores below classified as *low*. *Correct* classification indicates correspondence between high performance score and expert or low performance score and novice. Rating scale scores (performance scores) reflect the mean score (averaged between two raters).

Table 5
Accuracy between Performance Score and Experience Level (Simulation)

Performance Score (rating scale)	Experience Level (expert/novice)	Correct classification (mid-way point)	Correct classification (median-split)
4.29	Expert	Consistent	Consistent
3.57	Expert	Consistent	Consistent
2.08	Novice	Consistent	Consistent
2.22	Novice	Consistent	Consistent
2.00	Novice	Consistent	Consistent
1.29	Novice	Consistent	Consistent
1.22	Novice	Consistent	Consistent
2.86	Expert	Consistent	Inconsistent
3.79	Expert	Consistent	Consistent
3.86	Expert	Consistent	Consistent
2.93	Expert	Consistent	Inconsistent
3.72	Novice	Inconsistent	Inconsistent
4.29	Expert	Consistent	Consistent
4.29	Expert	Consistent	Consistent

Note. For mid-way point classification, performance scores above 2.5 classified as *high*; scores below as *low*. For median-split classification, performance scores above median score (3.248) classified as *high*; scores below classified as *low*. *Correct* classification indicates correspondence between high performance score and expert or low performance score and novice. Rating scale scores (performance scores) reflect the mean score (averaged between two raters).

Table 6

Descriptive Statistics for Top Five Most Intense Discrete Emotions over Time (Real Case)

	Top 5 Positive emotions		Top 5 Negative emotions	
	Discrete emotion	*M* (*SD*)	Discrete emotion	*M* (*SD*)
Time				
Before	Enjoyment	3.7 (0.9)	Anxious	2.4 (1.1)
	Relaxed	3.5 (1.0)	Afraid	2.1 (1.2)
	Hopeful	3.5 (1.1)	Frustrated	1.7 (1.0)
	Proud	3.4 (1.1)	Confused	1.5 (1.0)
	Compassionate	3.2 (1.3)	Disappointed	1.5 (0.8)
During	Enjoyment	3.6 (1.0)	Anxious	2.5 (1.5)
	Hopeful	3.3 (1.4)	Afraid	2.1 (1.4)
	Relaxed	3.2 (1.3)	Frustrated	1.7 (1.2)
	Proud	3.1 (1.4)	Disappointed	1.7 (1.2)
	Compassionate	3.1 (1.5)	Confused	1.6 (0.9)
After	Relaxed	3.6 (1.1)	Disappointed	1.7 (1.2)
	Enjoyment	3.5 (1.1)	Anxious	1.7 (0.9)
	Proud	3.4 (1.4)	Frustrated	1.6 (1.3)
	Relieved	3.2 (1.2)	Afraid	1.5 (0.9)
	Hopeful	3.1 (1.3)	Confused	1.5 (0.8)

Note. Given that the purpose is to present a representative range of the top five most intensely experienced emotions, happiness was removed to reduce redundancies as this emotion is highly correlated with enjoyment (*r*s = .71-.81).

Table 7
Descriptive Statistics for Top Five Most Intense Discrete Emotions over Time (Simulation)

	Top 5 Positive emotions		Top 5 Negative emotions	
	Discrete emotion	*M* (*SD*)	Discrete emotion	*M* (*SD*)
Time				
Before	Enjoyment	3.8 (0.8)	Anxious	2.0 (1.2)
	Relaxed	3.7 (1.2)	Confused	1.8 (1.1)
	Hopeful	3.6 (1.1)	Afraid	1.6 (1.1)
	Proud	3.1 (1.3)	Frustrated	1.1 (0.3)
	Compassionate	2.3 (1.2)	Disappointed	1.0 (0.0)
During	Enjoyment	3.9 (0.9)	Confused	1.9 (0.9)
	Hopeful	3.6 (0.8)	Anxious	1.5 (0.8)
	Proud	3.6 (0.9)	Afraid	1.4 (0.6)
	Relaxed	3.4 (1.3)	Frustrated	1.3 (0.6)
	Compassionate	2.1 (1.2)	Disappointed	1.2 (0.6)
After	Relaxed	3.9 (1.2)	Confused	1.4 (0.6)
	Enjoyment	3.5 (1.2)	Disappointed	1.2 (0.6)
	Proud	3.5 (1.2)	Frustrated	1.2 (0.6)
	Relieved	3.2 (1.2)	Afraid	1.1 (0.3)
	Hopeful	2.4 (1.5)	Anxious	1.1 (0.5)

Note. Given that the purpose is to present a representative range of the top five most intensely experienced emotions, happiness was removed to reduce redundancies as this emotion is highly correlated with enjoyment (*r*s = .71-.81).

Table 8

Correlations between Appraisals, Emotions, Regulation, Performance Variables (Real Case)

	1	2	3	4	5	6	7	8	9	10
1. Control appraisal										
2. Value appraisal	.39*									
3. Positive act. emotions	.56**	.44*								
4. Positive deact. emotions	.61**	.34	.79**							
5. Negative act. emotions	-.57**	.06	-.22	-.51*						
6. Negative deact. emotions	-.55**	-.09	-.46*	-.60**	.80**					
7. Electrodermal activity	.08	.16	-.03	.16	-.05	-.01				
8. Reappraisal strategies	.37+	.16	.15	.03	-.10	-.16	-.48*			
9. Suppression strategies	.04	-.16	.21	-.08	.24	.27	-.21	.19		
10. Technical performance	.52*	.24	.10	.22	-.43*	.05	.11	.39*	.22	
11. Self-efficacy future perf.	.74**	-.05	.21	.41*	-.80**	-.58**	-.03	.42*	-.02	.69**

Note. Emotional states are collapsed over time; +approaching significance, $p < .06$; * $p < .05$; ** $p < .01$; positive act. emotions = positive activating emotions; positive deact. emotions = positive deactivating emotions; negative act. emotions = negative activating emotions; negative deact. emotions = negative deactivating emotions; self-efficacy future perf. = self efficacy for future performance.

Table 9
Correlations between Appraisals, Emotions, Regulation, Performance Variables (Simulation)

	1	2	3	4	5	6	7	8	9	10
1. Control appraisal										
2. Value appraisal	.014									
3. Positive act. emotions	-.24	.01								
4. Positive deact. (relief)	-.52*	.012	.23							
5. Negative act. emotions	-.29	-.00	-.28	.58*						
6. Negative deact. emotions	-.29	.33	.07	.27	.70**					
7. Electrodermal activity	-.26	.06	.18	.54*	.15	.04				
8. Reappraisal strategies	-.32	.58*	.53*	.16	-.13	.07	.002			
9. Suppression strategies	-.00	-.40+	.15	-.24	-.14	-.31	.04	.07		
10. Technical performance	.50*	.08	-.41+	-.28	-.73**	-.60*	.13	-.06	-.01	
11. Self-efficacy future perf.	.37	.25	-.33	-.45	-.91**	-.59**	.06	.20	-.07	.83**

Note. Emotional states are collapsed over time; positive deactivating emotions sub-scale relaxation (over time) was not included as reliability was low, $\alpha < .50$; + approaching significance, $p < .082$; * $p < .05$; ** $p < .01$; positive act. emotions = positive activating emotions; positive deact. (relief) = positive deactivating emotion (relief only); negative act. emotions = negative activating emotions; negative deact. emotions = negative deactivating emotions; self-efficacy future perf. = self efficacy for future performance

Table 10

Descriptive Statistics for Appraisals, Emotions, Regulation, Performance (Real Case)

	Expert (*n* = 11)	Novice (*n* = 8)	*N* = 19
	M (*SD*)	*M* (*SD*)	α
Variables			
Control appraisal	4.55 (.55)	3.56 (.86)	.88
Value appraisal	4.36 (.86)	4.38 (.77)	.85
Positive activating emotions	3.13 (.87)	3.17 (.60)	.92
Positive deactivating emotions	3.15 (.80)	2.81 (.72)	.76
Negative activating emotions	1.23 (.30)	2.17 (.75)	.95
Negative deactivating emotions	1.23 (.37)	1.58 (.75)	.91
Reappraisal regulation strategy	5.20 (.89)	4.27 (2.17)	.90
Suppression regulation strategy	4.73 (1.91)	4.56 (1.55)	.93
Electrodermal activity	2.03 (1.85)	2.61 (1.85)	--
Technical performance	3.65 (.59)	1.75 (.80)	.97 (ICC=.95)
Self-efficacy future performance	93.64 (6.74)	48.75 (26.42)	--

Note. Emotional states are collapsed over time; dashes indicate single item scale/no reliability coefficient; the following discrete emotions were removed from sub-scales to improve reliability: disgust (all time points); ICC=Intra-class correlations (2-way random, average measures). Scale range is 1-5 for all variables, except post-task self-efficacy (0-100), and regulation strategy (1-7).

Table 11

Descriptive Statistics for Appraisals, Emotions, Regulation, Performance (Simulation)

	Expert (*n* = 8)	Novice (*n* = 6)	*N* = 14
	M (*SD*)	*M* (*SD*)	α
Variables			
Control appraisal	4.54 (.56)	3.72 (.44)	.83
Value appraisal	4.54 (.50)	4.33 (1.63)	.91
Positive activating emotions	3.29 (.81)	3.94 (.11)	.78
Positive deactivating (relaxation)	3.15 (.80)	2.81 (.72)	.46
Positive deactivating (relief)	1.54 (.56)	2.39 (1.14)	.73
Negative activating emotions	1.15 (.25)	1.78 (.62)	.81
Negative deactivating emotions	1.05 (.14)	1.43 (.45)	.71
Reappraisal regulation strategy	4.29 (1.24)	4.78 (1.33)	.87
Suppression regulation strategy	4.06 (1.24)	4.17 (1.68)	.93
Electrodermal activity (EDA)	2.11 (2.24)	3.85 (3.46)	*r* = .90
Technical performance	3.73(.58)	2.08 (.90)	.96 (ICC=.93)
Self-efficacy future performance	96.25 (10.60)	63.33 (29.43)	--

Note. Emotional states are collapsed over time; dashes indicate single item scale/no reliability coefficient; the following discrete emotions were removed from sub-scales to increase reliability: compassion (all time points), disgust (all time points), boredom (all time points), anger (pre), disappointment (pre), hopelessness (post), sadness (pre, post); given that relaxation and relief were poorly correlated at all time points, the positive deactivating emotion subscale was divided into separate discrete emotions; EDA data was correlated between devices on wrist and ankle; ICC=Intra-class correlations (2-way random, average measures); scale range is 1-5 for all variables, except post-task self-efficacy (0-100), and regulation strategy (1-7).

Table 12

Results of Qualitative Content Analysis of Interviews (Real Case)

Question	Category	Sample quotation	Overall pattern
What emotions did you experience? (real case)	Nervous/ anxious (n = 11)	"I was nervous going in" (about challenges, performance, and outcomes, such as tissue damage, vein quality, efficiency)	The most commonly reported positive emotions were happiness, pride, and calmness. The most commonly reported negative emotions were anxiety and frustration. Several participants also reported feeling neutral.
	Happy (n = 9)	"Once the vein was out I was happy" (about activity or success)	
	Proud (n = 8)	"Somewhat proud as well" (about demonstrating competence or reaction to success)	
	Calm (n = 8)	"Pretty calm most of the time" (about activity and outcome)	
	Neutral (n = 5)	"I think I felt pretty neutral"	
	Relieved (n = 4)	"When [supervisor] took over, that was a relief" (about success or reduced stressors)	

Question	Category	Sample quotation	Overall pattern
	Frustrated (n = 3)	"Frustrating when you get faced with a hurdle" (about challenges, such as bleeding, fragile tissue or performance/outcomes)	
	Confused (n = 1)	"A little bit confused as to what I'm supposed to do" (about procedural challenge or lack of task clarity)	
	Grateful (n = 1)	"I am grateful [for opportunity, for assistance]" (about help or opportunity)	
	Curious (n = 1)	"At the beginning, [I was] curious to find the vein" (about the task, solving problems)	
	Disappointed (n = 1)	"[I] was a tiny bit disappointed" (about performance or outcomes)	
What factors contributed to these emotions? (real case)	Experience/ knowledge/ skills (n = 14)	High: "I think once the vein was out and I relaxed because I knew what I was doing" Low: "I think with assistance I feel much more confident that I could do this procedure [than] when I'm alone" (due to lack of experience)	Higher levels of experience/skill level often associated with positive emotions. Lower levels of experience often associated with more negative emotions.

Question	Category	Sample quotation	Overall pattern
	Performance (*n* = 13)	High: "I think it was positive because that vein was taken out in two seconds" Low: "I think I could have done a lot better"	Expectations or reactions to successes often associated with positive emotions. Concerns or reactions to failures often associated with negative emotions.
	Task importance/ interest (*n* = 12)	Positive emotion: "I'm doing the thing I like to do" Negative emotion: "I believe that if I do care, I have to be stressed out for my first time."	Higher task importance associated with both positive (enjoyment) and negative emotions (anxiety). Lower task importance in some cases associated with lower emotional intensity.
	Difficulty/ demands of procedure (*n* = 12)	High: "Well, everyone made it look so easy. But when you get in there everything is so slippery" Low: "Pretty calm most of the time because it wasn't that bad that you have to take four sections of vein"	High task difficulty often associated with negative emotions. Low task difficulty often associated with positive emotions.
	Team dynamics (*n* = 10)	Positive: "The resident, he was very helpful" Negative: "Some of the surgeons they make you more nervous, when you do something wrong"	Positive team dynamics often associated with positive emotions. Negative team dynamics often associated with negative emotions.

Question	Category	Sample quotation	Overall pattern
	Time constraints/ pressure (*n* = 8)	High: "If they are moving really fast and they need the vein right away because it's an emergency" Low: "there was no rush, so it was pretty calm"	High time constraints often associated with negative emotions. Low pressure often associated with positive emotions.
	Task initiation (*n* = 3)	"The first suture, the first bite… when I was starting the whole operation"	Some participants mentioned initiating the task as stressful.
What regulation strategies do you use to help you manage your emotions (general surgery)?	Controlling expression/ behavior (*n* = 9)	Speech Modulation: "I'd probably talk less" (*n* = 4) Suppression: "I just internalize all those emotions" (*n* = 5) Breathing techniques: "I take a deep breath and move on" (*n* = 3)	The most common emotion regulation strategies were controlling expression, help seeking, reappraisal, and distraction, although several other strategies were mentioned as well. Strategies were more likely to be a response to negative emotions
	Help-seeking (*n* = 6)	"When I don't think I have other alternatives, I would just seek help"	
	Reappraisal (*n* = 5)	"Like thinking positively about the situation itself. So you can change the way you think"	

Question	Category	Sample quotation	Overall pattern
	Distraction (*n* = 4)	"I'll look around and try to think about something alternative"	
	Concentration (*n* = 3)	"Trying to focus on one step at a time"	
	Acceptance/ mindfulness (*n* = 3)	"I wouldn't want to change even the negative feelings"	
	Problem-solving (*n* = 3)	"I don't freak out, I think how to fix it"	
	Competence increase (*n* = 1)	"Well try to practice more and try not to make mistakes if you make mistakes, so next time you don't feel this"	
	Savoring (*n* = 1)	"When I had a positive emotion, I clinched it"	
	Regulating others (*n* = 3)	Call attention from others: "If I am angry, I will shout" Motivating others: "I try to motivate the residents, like 'Oh, it's good, his heart is good, functioning well' "	

Note. Strategies include specific for case completed and more general strategies used for other cases within the operating room. Total *N* = 16 as 3 participants could not fully complete interview due to time constraints (e.g., needed to return to operating room or clinic). Parentheses next to emotion quotations provide examples of object-focus. Sample quotations and corresponding content in parentheses are intended to provide examples and are not exhaustive. *Real case* indicates that responses are specific to the real case recently completed. *General surgery* indicates that responses refer to experiences recalled in surgical settings more broadly.

Table 13
Results of Qualitative Content Analysis of Interviews (Simulation)

Question	Category	Sample quotation	Overall pattern
What emotions did you experience? (simulation)	Curious (n = 8)	"I was pretty curious to see what it was going to be like" (about the task, solving problems)	The most commonly reported positive emotions for the simulation were curiosity and happiness. The most commonly reported negative emotion was anxiety.
	Happy/excited/ enjoyed (n = 8)	"I felt quite happy and enjoyed it" (about activity or success)	
	Relaxed (n = 4)	"You know it's not the real thing so I'm kind of more relaxed" (about activity and performance)	
	Afraid/anxiety (n = 4)	"I was afraid of damaging something" (about challenges, such as fragile tissue, or performance)	
	Hopeless (n = 2)	"I really felt a bit hopeless, helpless" (about performance, about knowing procedures)	
	Confused (n = 2)	"A little bit of a confusion" (about procedural challenges or lack of task clarity)	

Question	Category	Sample quotation	Overall pattern
	Proud/confident (n = 2)	"During the procedure I felt confident" (about demonstrating competence or reaction to success)	
	Surprised (n = 2)	"I was quite surprised at the fidelity of it [simulation model]" (about unexpected event or outcome)	
	Disappointment (n = 1)	"Disappointment because I didn't know how to do it" (about performance)	
	Relief (n = 1)	"Afterwards, a bit of relief that I didn't mess up" (about success or reduced stressors)	
	Neutral (n = 1)	"I would say neutral"	
What factors contributed to these emotions? (simulation)	Task importance/ interest (n = 6)	"I wanted to learn something"	The most commonly reported antecedents of a negative emotion in the simulation task were concerns over performance/evaluation or lack of experience. The most commonly reported antecedents of positive emotions in the simulation were task importance/interest and novelty.

Question	Category	Sample quotation	Overall pattern
	Novelty ($n = 5$)	"I was pretty curious to see what it was going to be like"	
	Performance/ evaluation ($n = 5$)	"I was scared of doing something wrong"	
	Experience/ knowledge/skills ($n = 4$)	"It was the first time I ever did it"	
	Fidelity ($n = 4$)	"I was quite surprised at the fidelity of it"	
	Difficulty/ demands of procedure ($n = 3$)	"Here tissue was flimsy and I was afraid of damaging something"	

Question	Category	Sample quotation	Overall pattern
How did you feel compared to a real case? (simulation versus real case)	Consequences/ time pressure (*n* = 9)	"If it's an emergency case than you're more stressed [in real case]"	Negative emotion typically reported to be more intense in real case (e.g., stress). Differences in the nature of the task included greater consequences, time pressure, and desire to demonstrate competence for real case. However, some participants noted that stress could be higher in simulation if observed or evaluated closely. Some participants felt there was no difference in emotion. This varied based on experience level and nature of real case compared to (e.g., emergency versus routine).
	Evaluation (*n* = 4)	"I would like to impress my superiors in the OR [real case] more than a simulation" (N=2) If a staff is looking at you during the simulation, I think it will be more stressful than in a real OR" (*n* = 1)	
	Emotion intensity (*n* = 7)	"Oh I would feel a lot more scared and even a lot more excited [in real case]"	

Question	Category	Sample quotation	Overall pattern
	No difference ($n = 4$)	"No difference really"	
What is the optimal emotional state for you to perform well? (general surgery)	(some) Stress ($n = 5$)	"A little bit of stress, a little bit of pressure feeling… enough to make you focus"	Several participants noted that states of calmness and enjoyment/ excitement were important. Several noted that a small amount of anxiety helped to focus attention and improve performance. Several participants described multiple emotional states (complex/ blended emotional states).
	Enjoyment /excitement ($n = 4$)	"Really important to be excited about it, to go to it"	
	Calmness ($n = 2$)	"Ideal state, I think, is calm"	
	Irrelevant ($n = 2$)	"Well I think if I have to do it, I have to do it. Emotion has nothing to do with the task"	

Question	Category	Sample quotation	Overall pattern
What regulation strategies do you use to help you manage your emotions (general surgery)?	Reappraisal ($n = 3$)	"You need to put a positive spin on things"	Participants commonly mentioned using reappraisal and help-seeking strategies in surgical settings to manage negative emotions, although other strategies, such as concentration and controlling expression were mentioned as well.
	Help-seeking ($n = 3$)	"If I don't know what to do then I ask someone"	
	Concentration ($n = 2$)	"I try to focus more on what I'm doing and I don't think of anything else during the procedure so just to stay focused"	
	Controlling expression ($n = 2$)	Behavioral control: "by behaving professionally and demonstrating the attitude that you'd like to see people behave in" Breathing techniques: "try to take a deep breath, a few deep breaths and hopefully that will try to…get back in a less anxious state"	
	Problem-solving ($n = 1$)	"I try to keep them in check, see what I can do to make the situation better"	

Question	Category	Sample quotation	Overall pattern
	Competence Increase (*n* = 1)	"I just try to inform myself more about the case and try to know more"	
What aspects of the model were realistic? (simulation)	Anatomy (*n* = 8)	"The position of the vein was close anatomically to what we expect it to"	Participants most commonly mentioned the anatomy, tissue, and procedural steps as being realistic
	Tissue (*n* = 5)	"I think the quality of the stuff you're using mimics some of the real tissues"	
	Procedural steps/feeling of procedure (*n* = 3)	"The steps, the process"	
	Suturing (*n* = 1)	"I think the closing was good"	
	Instrument handling (*n* = 1)	"How to handle the instruments"	
	Branches (*n* = 1)	"[the] Branches"	

Question	Category	Sample quotation	Overall pattern
What aspects of the model could be improved or were unrealistic? (simulation)	Branches (*n* = 7)	"The branches were less realistic"	Most participants mentioned that the branches could be improved. Other commonly noted areas for improvement included the tissue texture, vein size, and increased complexity.
	Tissue texture (*n* = 6)	"How the subcutaneous tissues can be; if it can be more dense then we can put the sutures in it"	
	Vein (*n* = 5)	"The size of the vein is larger than the real-life"	
	Increase complexity (*n* = 4)	There's less of the chaos that happens in the OR [real case] or the complexity… that you don't have here"	
	Stability (*n* = 1)	"It needs to be more stable"	
What or who do you think this model is most useful for? (simulation)	Medical students (*n* = 8)	"Medical student, basic surgical skills"	
	Junior residents (*n* = 5)	"Junior residents. First and second year residents I think it would be a useful tool"	

Question	Category	Sample quotation	Overall pattern
	Provides practice for those with more experience (*n* = 3)	"For people that may have even had some experience with it, and want to look at a more challenging procedure"	
	Learning technical procedures/steps (*n* = 4)	"Vein harvesting… how to tie vascular"	
	Instrument handling (*n* = 4)	"You can learn how to use the instruments while you're doing the dissection"	
	Tissue Handling (*n* = 2)	"Tissue handling skill because if you're rough than you're going to rip everything, nothing is going to close."	
	Troubleshooting/ taking time (*n* = 2)	"In real case it is time limited" "for troubleshooting"	
	Build confidence/stay calm (*n* = 2)	"What do you do if you encounter this, and you need a calm situation" "this is what builds confidence"	

Note. Parentheses next to emotion quotations provide examples of object-focus. Sample quotations and corresponding content in parentheses are intended to provide examples and are not exhaustive. *Simulation* indicates that responses are specific to the simulation recently completed. *Simulation versus real case* indicates that responses involve a comparison between the two conditions. *General surgery* indicates that responses refer to experiences recalled in surgical settings more broadly (several participants felt that they did not use sufficient regulation strategies during simulation to elaborate their responses so it was expanded to general surgery).

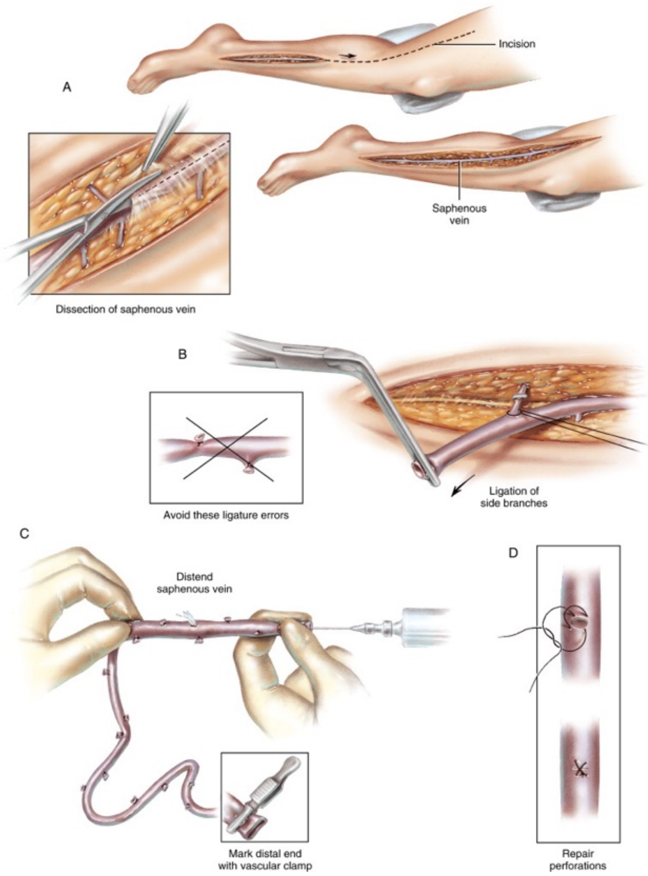


Figure 5. Open saphenous vein harvest procedure.

Figure from "Coronary Artery Bypass Graft," by D. B Doty & J. R. Doty (Eds.), *Cardiac surgery: Operative technique (2nd Ed.)* (p. 397), 2012, London, England: Elsivier. Copyright 2012 by Elsivier. Reprinted with permission.

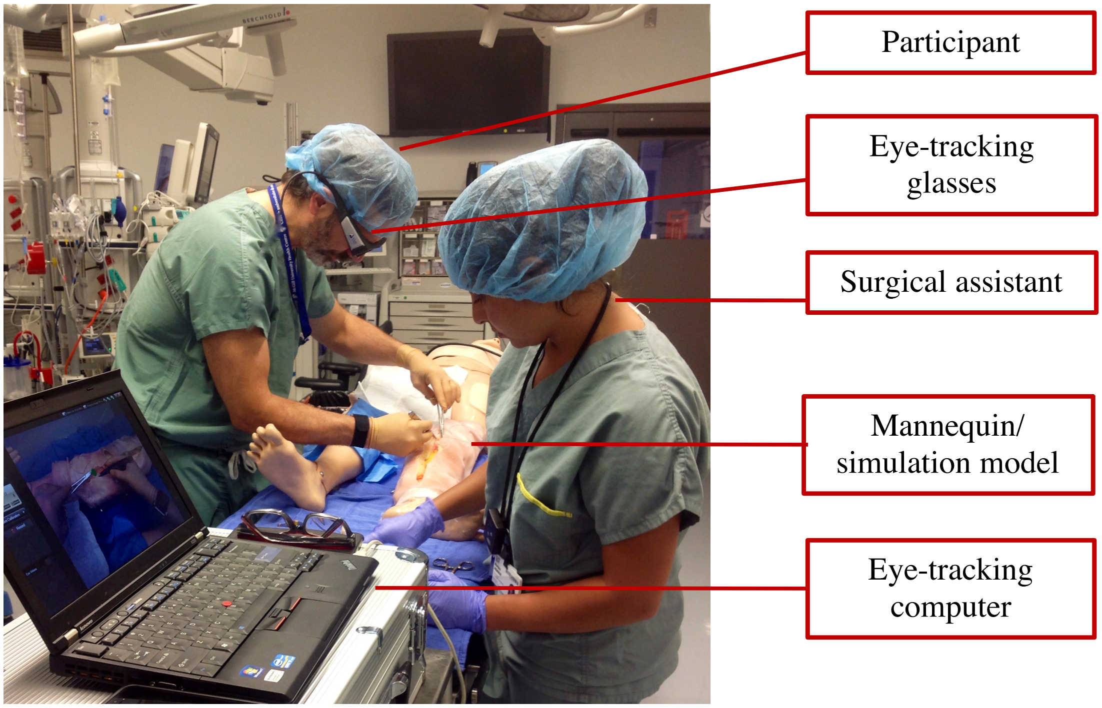


Figure 6. Experimental and surgical set-up in simulation case.

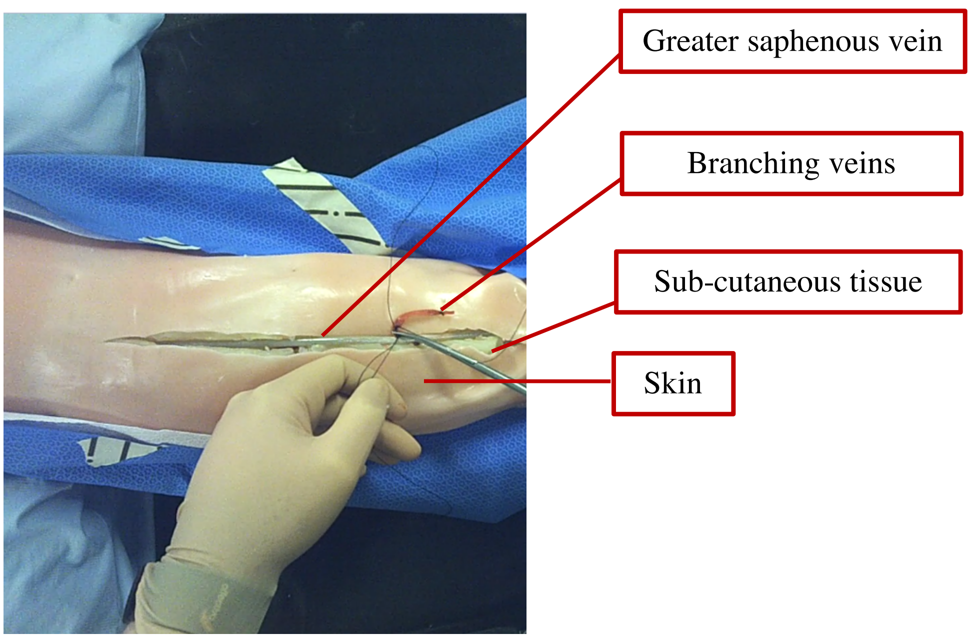


Figure 7. Close-up of simulation vein model

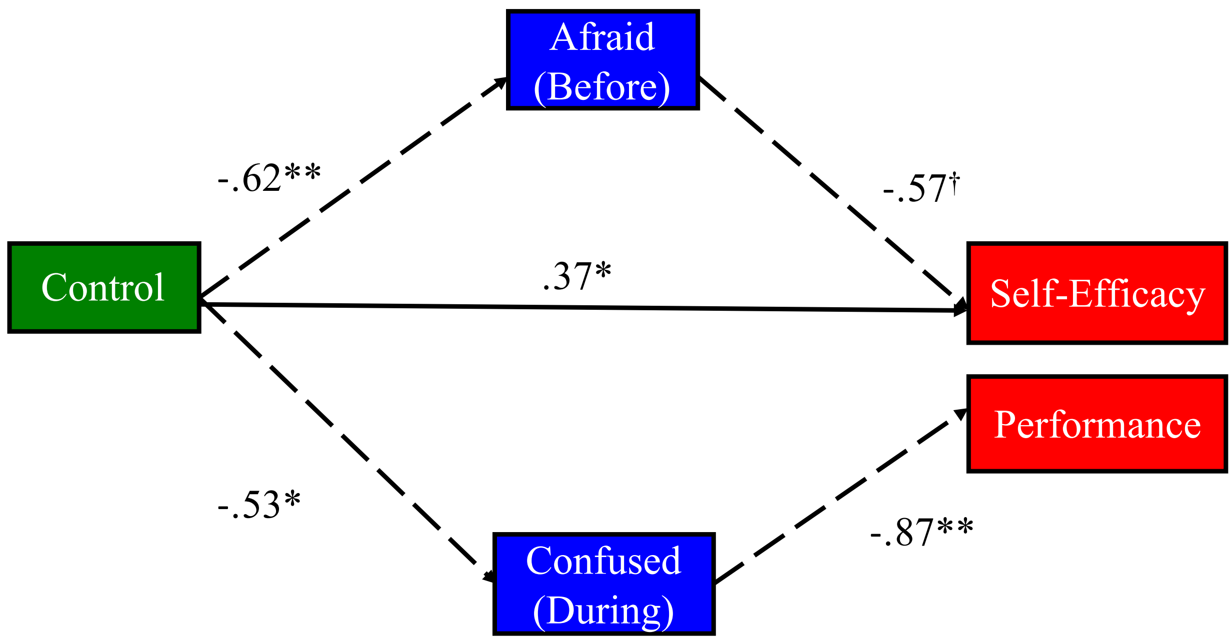


Figure 8. Mediation model using Hayes & Preacher (bootstrap method)
†p = .05; *p < .05; **p < .01.

Appendix A
Scale/Questionnaire Items

Prior experience:
How many times have you observed this surgical procedure in the operating room?

○	○	○	○	○
0	1-5	6-10	11-15	21+

How many times have you conducted this surgical procedure in the operating room?

○	○	○	○	○
0	1-5	6-10	11-15	21+

Emotional State Scale:

	1 Not at all	2 Very little	3 Moderate	4 Strong	5 Very strong
Confused	○	○	○	○	○
Bored	○	○	○	○	○
Hopeful	○	○	○	○	○
Proud	○	○	○	○	○
Sad	○	○	○	○	○
Anxious	○	○	○	○	○
Frustrated	○	○	○	○	○
Happy	○	○	○	○	○
Hopeless	○	○	○	○	○
Enjoyment	○	○	○	○	○
Ashamed	○	○	○	○	○
Neutral	○	○	○	○	○
Compassionate	○	○	○	○	○
Surprised	○	○	○	○	○
Curious	○	○	○	○	○
Afraid	○	○	○	○	○
Grateful	○	○	○	○	○
Disappointed	○	○	○	○	○
Relieved	○	○	○	○	○
Angry	○	○	○	○	○
Relaxed	○	○	○	○	○
Disgusted	○	○	○	○	○

Control Appraisal Scale:
Within this environment, I feel I have control over:

	1 strongly disagree	2 somewhat disagree	3 neutral	4 somewhat agree	5 strongly agree
My surgical performance	○	○	○	○	○
The patient's outcome	○	○	○	○	○
What I am expected to do	○	○	○	○	○
Surgical assistance	○	○	○	○	○

Value Appraisal Scale:
Conducting this surgical procedure is:

	1 strongly disagree	2 somewhat disagree	3 neutral	4 somewhat agree	5 strongly agree
Useful to me	○	○	○	○	○
Interesting to me	○	○	○	○	○
Important to me	○	○	○	○	○

Self-efficacy Item:
Rate how confident you are that you can effectively conduct this surgical procedure

○	○	○	○	○	○	○	○	○	○	○
0	10	20	30	40	50	60	70	80	90	100
not at all certain					moderately certain					highly certain

Proficiency item:
Rate your level of proficiency in conducting this surgical procedure.

○	○	○	○	○
1	2	3	4	5
poor	fair	average	good	very good

Satisfaction Item:
I am satisfied with my performance during the surgical procedure

○	○	○	○	○
1	2	3	4	5
strongly disagree	somewhat disagree	neutral	somewhat agree	strongly agree

Emotion Regulation Scale (Gross & John, 2003):

○	○	○	○	○	○	○
1	2	3	4	5	6	7
strongly disagree			neutral			strongly agree

1. When I wanted to feel more positive emotion (such as joy or amusement), I changed what was thinking about.
2. I kept my emotions to myself.
3. When I wanted to feel less negative emotion (such as sadness or anger), I changed what I was thinking about.
4. When I was feeling positive emotions, I was careful not to express them.
5. When I was faced with a stressful situation, I made myself think about it in a way that helped me stay calm.
6. I controlled my emotions by not expressing them.
7. When I wanted to feel more positive emotion, I changed the way I was thinking about the situation.
8. I controlled my emotions by changing the way I thought about the situation I was in.
9. When I was feeling negative emotions, I made sure not to express them.
10. When I wanted to feel less negative emotion, I changed the way I was thinking about the situation

Simulation Model Feedback Scale:

I found the simulation vein model realistic

○	○	○	○	○
1	2	3	4	5
strongly disagree	somewhat disagree	neutral	somewhat agree	strongly agree

The simulation vein model was useful for practicing my technique.

○	○	○	○	○
1	2	3	4	5
strongly disagree	somewhat disagree	neutral	somewhat agree	strongly agree

What did you like about the simulator?

What (if anything) would you like to see improved?

Appendix B
Global Rating Scale for Technical Performance

GLOBAL RATING SCALE OF OPERATIVE PERFORMANCE

Please circle the number corresponding to the candidate's performance in each category, **irrespective of training level.**

	1	2	3	4	5
Respect for Tissue:	Frequently used unnecessary force on tissue or caused damage by inappropriate use of instruments		Careful handling of tissue but occasionally caused inadvertent damage		Consistently handled tissues appropriately with minimal damage
Time and Motion:	Many unnecessary moves		Efficient time/motion but some unnecessary moves		Clear economy of movement and maximum efficiency
Instrument Handling:	Repeatedly makes tentative or awkward moves with instruments by inappropriate use of instruments		Competent use of instruments but occasionally appeared stiff or awkward		Fluid moves with instruments and no awkwardness
Knowledge of Instrument:	Frequently asked for wrong instrument or used inappropriate instrument		Knew names of most instruments and used appropriate instrument		Obviously familiar with the instruments and their names
Flow of Operation:	Frequently stopped operating and seemed unsure of next move		Demonstrated some forward planning with reasonable progression of procedure		Obviously planned course of operation with effortless flow from one move to the next
Use of Assistants:	Consistently placed assistants poorly or failed to use assistants		Appropriate use of assistants most of the time		Strategically used assistants to the best advantage at all times
Knowledge of Specific Procedure:	Deficient knowledge. Needed specific instruction at most steps		Knew all important steps of operation		Demonstrated familiarity with all aspects of operation

OVERALL ON THIS TASK, SHOULD THE CANDIDATE: FAIL PASS

Image from "Testing Technical Skill Via an Innovative 'Bench Station' Examination," by R. Reznick, G. Regehr, H. MacRae, J. Martin, & W. McCulloch, 1997, *The American Journal of Surgery, 173*, 229.

Appendix C

Interview Question Guide (Real Case)

1. What emotions were you experiencing before/during/after the surgical procedure?
    - Follow-up prompts: Were you experiencing any positive emotions? Were you experience any negative emotions? Which emotions in particular? How did you feel? Did your emotions change over time?
2. What caused you to feel that way?
    - Follow prompts: What factors contributed to these emotions? What led to that you to experience that emotion? Who or what was the emotion directed towards?
3. Did you find the surgery stressful?
    - Follow-up prompts: were you anxious at any point? Were you relieved at any point?
      Which technical components of the surgical procedure did you find more stressful?
4. Do you think you were you able to effectively monitor and regulate your emotions during the surgery?
    - Follow-up prompts: How did you do this? What did you do to regulate your emotions? What do you typically do or have you done in other cases within the operating room?
    - Follow prompts: are they similar or different? Were you more or less anxious?
5. Were the emotions you experienced typical for you for this kind of procedure?
6. Did you find the emotions listed on the scale to be appropriate and relevant?
    - Follow prompts: were there any irrelevant emotions? Were there any other emotions you would add?

Appendix D

Interview Question Guide (Simulation)

1. What emotions were you experiencing before/during/after the simulation procedure?
    - Follow-up prompts: Were you experiencing any positive emotions? Were you experience any negative emotions? Which emotions in particular? How did you feel? Did your emotions change over time?
2. What caused you to feel that way?
    - Follow prompts: What factors contributed to these emotions? What led to that you to experience that emotion? Who or what was the emotion directed towards?
7. Do you think you were you able to effectively monitor and regulate your emotions during the procedure?
    - Follow-up: How did you do this? What did you do to regulate your emotions? What do you typically do or have you done in other cases within the operating room?
3. How did your experience during the simulation procedure compare to a real case within the operating room?
    - Follow-up prompts: were the emotions you experienced similar or different from what you would experience for a real case? How so?
4. What is the optimal emotional state for you to be in to perform well during surgical procedures?
    - Follow-up prompts: What kinds of emotions help you to feel like to you can perform at an optimal level? Is there a particular emotional state that helps you to perform well?
5. How do you manage your feelings if you're not in this state?
    - Follow-up prompt: Do you use any strategies to monitor or regulate your emotions?
6. Do you think the simulation model was realistic?
    - Follow-up prompts: Which aspects? What would you change, if anything? How could it be improved?
7. Do you think the simulation model provides a useful training opportunity?
    - Follow-up prompts: When do think this model would be most appropriate to use? Who would benefit from using this model?
8. Did you take this task seriously?
    - Follow-up prompt: Did you complete the procedure as fast or with the same amount of care as real case?

Appendix E
Interview Coding Scheme

Definitions, Examples & Instructions for Interview Coding

Emotion Code Definition: an emotion is an affective state that has a specific object-focus (directed towards an agent or event) and is shorter in duration than a mood or general disposition. Typically referred to as a feeling that may also be described in combination with cognitive (thinking), motivational (goal-directed behavior), physiological (e.g., sweating, heartrate), and behavioral (e.g., facial expression, body language, vocal tone) processes.
Examples of Discrete Emotion States: angry, sad, excited, happy, proud, confused, frustrated, disappointed, hopeless, bored, relaxed, relieved, anxious (see complete list of emotions in scale)
Examples of Complex/Blended Emotion States: good, bad (in most cases follow-up prompts will help to identify discrete states)
Emotion Code Instructions: Code each discrete emotion state described. For each emotion:

- Indicate the object focus or what the emotion is directed towards
- (e.g., directed toward self, another person, event).
- Indicate point in time that the emotion was experienced (i.e., before, during, after task)
- Indicate whether emotion is positive, negative, or neutral (neutral, surprise)
- Indicate intensity and duration of emotion if described (e.g., feeling *very* angry, feeling *a little* anxious for a *short time*)

Antecedent Code Definition: an antecedent refers to factors or events that caused, contributed, or linked to the activation, onset, or intensity of an emotion.
Examples of Antecedents: time pressure, team dynamics, task difficulty/complexity, experience/skill level, task novelty, task fidelity, task importance, performance/outcome.
Antecedent Code Instructions: Code each type of antecedent described. For each antecedent:

- Indicate whether it was high (e.g., a great deal of time pressure) or low amount (e.g., no time pressures) or whether it was appraised as a positive event (good team dynamics) or negative event (e.g., problematic team dynamics)
- Indicate whether it was linked to positive or negative emotion

Emotion Regulation Definition: emotion regulation refers to managing or controlling emotions to achieve a more optimal emotional state (e.g., attempts to feel less anxious). Regulation processes may include monitoring emotional states, as well as using strategies to change, modify, or prolong a current or future emotional state.
Examples of Regulation Strategies: reappraisal (change way event is thought about), control expression (supress facial expression, change vocal tone), help-seeking (ask someone for assistance), problem-solving (seek solutions to solve issue), distraction (think about something else), concentrate (focus attention), acceptance (non-judgemental recognition of emotion), savor (mindful appreciation of emotion), increase competence (e.g., practice), regulate others.
Emotion Regulation Code Instructions: Code each type of regulation strategy described. For each description of regulation:

- Indicate purpose of regulation (i.e., whether it was intended to increase/decrease intensity or extend/shorten duration of positive or negative emotion)
- Whether it was deliberate (e.g., effortful or intentional effort) or automated (habitual or occurred naturally without thought)

Appendix F
Supplemental analyses:
Relations between discrete emotions, performance, and self-efficacy

Correlation analysis (two-tailed tests) were conducted to select the discrete emotions that demonstrated the strongest relationship to performance and self-efficacy. In terms of discrete emotions experienced before the task, results revealed that pride was positively related to both performance (r = .47, p < .05) and self-efficacy (r = .54, p < .01) and relaxation was also positively related to self-efficacy (r = .71, p < .01). Anxiety was negatively related to both performance (r = -.53, p < .05) and self-efficacy (r = -.80, p < .01). Fear was also negatively related to both performance (r = -.61, p < .01) and self-efficacy, (r = -.83, p < .01). Sadness was negatively related to self-efficacy (r = -.46, p < .05), as was hopelessness (r = -.58, p < .01).

In terms of discrete emotions experienced during the task, relaxation was positively related to both performance (r = .54, p < .05) and self-efficacy (r = .77, p < .05). Happiness was also positively related to self-efficacy (r =.59, p < .01). Confusion was negatively related to performance (r = -.73, p < .01) and self-efficacy (r = -.74, p < .01). Fear was also negatively related to both performance (r = -.63, p < .01) and self-efficacy (r = -.83, p < .05). Surprise was also negatively related to both performance (r = -.49, p < .05) and self-efficacy (r = -.65, p < .01). Sadness was negatively related to self-efficacy (r =-.57, p < .05) as was anxiety (r =-.76, p < .01), frustration (r =-.71, p < .01), disappointment (r = -.78, p < .01), and hopelessness (r = -.52, p < .05).

Bridging Text

In Chapter 3, I presented an empirical study that examined the nature of emotions and their relations to appraisals, performance, and task outcomes, for authentic training conditions situated within the domain of surgery. The results provided strong support for the control-value theory, particularly in the real case within the operating room. Findings revealed that negative activating emotions, which are particularly detrimental for performance and task outcomes, are highest among novices. However, the findings suggest that higher levels of perceived control before the task can help to reduce the intensity of negative emotions and the use of reappraisal regulation strategies during the task can improve performance. Qualitative analyses identified additional constructs for future empirical work, such as context-specific antecedents that may contribute to the onset of emotions as well as a diverse range of regulation strategies.

Whereas Chapter 3 focused on a surgical environment, Chapter 4 expands on these findings through a broader lens in which emotions are examined across three authentic medical learning environments representing clinical reasoning, emergency care, and surgical skills. Consistent with Chapter 3, this empirical study employs both quantitative and qualitative methods to triangulate findings. In addition, Chapter 4 continues to focus on state rather than trait emotions. The goal is to address limitations in Chapter 3 to test whether the findings linking appraisals, emotions, and performance are robust and generalizable across medical education environments. This study also builds from the recommendation in Chapter 3 manuscript to re-examine relations between emotions and whether different profiles of emotional experiences occur outside of the standard grouping of valence X activation. At the same time, this study takes a closer examination of the assumptions underlying the measurement of emotions by assessing the validity of questionnaire measures and the alignment between multiple measures (e.g., retrospective interview versus self-report scale). In doing so, it offers a valuable review on the use of single-item measures of emotions and addresses commonly cited critiques of self-report.

Chapter 4: Manuscript 3

Emotions in medical education: Examining the validity of a self-report emotions scale and the generalizability of control-value theory across authentic medical learning environments

Earlier versions of this manuscript were presented at the following conferences:

Duffy, M. C., Lajoie, S. P., Pekrun, R., & Lachapelle, K. (2016, April). *Mapping emotions to appraisals and performance in authentic medical learning environments*. Paper presented at the conference of the American Educational Research Association, Washington, DC.

Duffy, M. C., Lajoie, S., Jarrell, A., Pekrun, R., Azevedo, R., & Lachapelle, K. (2015, April). *Emotions in medical education: Developing and testing a self-report emotions scale across medical learning environments*. Paper presented at the conference of the American Educational Research Association, Chicago, IL.

Abstract

To advance research on emotions in medical education, there is a need to develop practical and context-relevant measures of emotion and to test the applicability of well-established emotion theories within medical education. The purpose of this research was to (1) develop and examine the initial validity of a self-report scale (the Medical Education Emotion Scale, MEES) designed to measure the unique range of emotions activated within medical education; (2) test the applicability of control-value theory of achievement emotions within medical education; and, (3) to identify areas for further inquiry in emotion theory and measurement in medical education. In Study 1, retrospective and cognitive interviews were conducted with a sample of medical trainees (N = 15) to triangulate emotion data, examine participant response process, and identify distal antecedents of emotions. In Study 2, medical trainees' (N = 60) emotions, appraisals, performance, and post-task self-efficacy were examined using questionnaires and context-specific measures of performance for three learning environments: (1) a computer-based learning environment designed to scaffold diagnostic reasoning skills; (2) a surgical environment designed to scaffold technical skills; and, (3) a simulation environment designed to scaffold critical care procedural skills. Results from Study 1 demonstrated that the self-report scale captured an appropriate range of emotions and that there was alignment between scale responses and interview responses. Results from Study 2 revealed that emotions measured using the MEES were correlated to theoretically relevant constructs (e.g., appraisals, performance) consistent with control-value theory. Results also revealed that profiles and the component structure of the scale were largely consistent with emotion theory. Findings from this research provide support for the generalizability of control-value theory and initial validation of the MEES. Recommendations for advancing emotion theory and measurement in future work are discussed.

Keywords: emotion; measurement; validation; medical education

Introduction

The field of medical education has largely attended to the role of cognition and metacognition. However, recent calls to advance our understanding of the role of emotions in learning (e.g., Artino & Durning, 2011; LeBlanc et al., 2015; McConnell & Eva, 2012) has led to deliberate efforts to place emotions at the forefront of empirical investigations rather than the periphery. To address these calls, there is a need to develop valid and practical measures of emotions that can be used across a range of medical education environments. Existing emotion questionnaires are not tailored to authentic medical education environments, which may include technology-rich environments, simulation environments, and clinical settings. As such, the primary purpose of this research was to develop and examine the validity of a self-report scale designed to capture the unique range of emotions experienced in medical education. At the same time, this research provides an opportunity to test the generalizability and utility of control-value theory of achievement emotions across a variety of medical education tasks. To *systematically* study the role of emotions in medical education, research should be grounded in well-established emotion theories that help to identify relevant constructs and explain complex relations. Although researchers have advocated for the use of control-value theory of achievement emotions (Pekrun, 2006; Pekrun & Perry, 2014) in medical education (see Artino et al., 2012), limited research has directly tested its applicability within this domain.

The present study focuses on authentic learning environments (problem-based tasks within real-world training or simulation counterparts) to test the generalizability of control-value theory outside of the classroom (cf. Artino et al., 2010), and examine the validity of a self-report scale designed to measure emotions in medical education. This research also contributes to the literature by presenting advances in emotion measurement and analytics. Specifically, this research employs several techniques to provide richer insight into the nature of emotions through: triangulation of multiple data channels (interviews and questionnaires), quantitative and qualitative analyses (inferential statistics, content analyses), nomothetic (i.e., group-centered) and idiographic (i.e., person-centered) analyses, as well as confirmatory (hierarchical regression) and exploratory (cluster, principal component) analyses. The results from this work will facilitate future investigation of emotions in medical education by assessing the validity of emotion measures and theories, as well as identifying areas requiring further empirical scrutiny. In the following sections, we describe the theoretical framework underlying this work, followed by a critical review of existing self-report scales used to examine emotions.

Theoretical Framework

To accurately detect emotions, measures must be carefully considered in terms of their alignment to subscribed theories and definitions (Artino & Naismith, 2015). The framework underlying our definition and measurement of emotions is control-value theory (CVT) of achievement emotions (Pekrun, 2006; Pekrun & Perry, 2014). According to this theory, cognitive appraisals of control and value lie at the heart of emotion-generation (Pekrun, 2006; Pekrun & Perry, 2014). In other words, individuals will experience specific emotions depending on the extent to which they feel they have agency over the activities or outcomes (i.e. subjective control appraisal) and depending on the extent to which they consider the task to be important to them (i.e., subjective value appraisal, Pekrun, et al., 2007; Pekrun, 2006). For instance, an individual who highly values a task but does not feel personally in control of the outcome should experience anxiety. In contrast, high task value and high levels of control should lead to feelings of enjoyment. These subjective task appraisals, that are considered proximal antecedents of emotion experiences, are influenced by distal antecedents including learner characteristics as well as factors embedded in the task design and broader socio-cultural environments (e.g., nature of feedback, peer support, task complexity). Once emotions are activated, they can influence performance through the mediated role of cognition, motivation, and metacognition (Pekrun et al., 2006). This theory is well suited to medical education given that it was developed specifically for achievement settings (i.e., environments that involve learning and performance tasks with success or failure outcomes, Pekrun & Linnenbrink-Garcia, 2014) and delineates constructs and pathways that can be used to examine concurrent validity (e.g., relations between control/value appraisals and emotions) and predictive validity (e.g., relations between emotions and task outcomes).

Regarding our definition of emotions, we consider emotions to fall within the broader concept of affect but, consistent with Artino and Naismith (2015), we treat them as distinct from other affective phenomena, such as moods, in that they are more intense, have a clearer object-focus, a more salient cause, and are experienced for a shorter duration (Scherer, 2005; Shuman & Scherer, 2014). We also endorse a multi-componential view of emotions in that each emotion is composed of coordinated sets of psychological sub-systems, including cognitive, motivational, expressive, physiological, and subjective feeling components (Ellsworth, 2013; Moors, 2014;

Moors, Ellsworth, Scherer, & Frijda, 2013). We consider emotions to be continuous and recursive episodes (Scherer, 2005; 2009; Shuman & Scherer, 2014), such that at any point in time the psychological subsystems are activated to varying degrees, which can give rise to changes in emotional states over time (Ellsworth, 2013; Moors et al., 2013). In the following section we discuss self-report scales—the most commonly used method of measuring emotions.

Existing Emotion Scales

Self-report measures of emotions rely on individuals' interpretations and reports of personal emotional states, often collected through questionnaires or scales. Although, some researchers have used other methods to measure emotions, such as physiological arousal or behavioral cues, self-report—and in particular questionnaires—continue to remain one of the most popular methods to capture meaningful emotional experiences due to their relative efficiency and ability to capture diverse states (Duffy, Lajoie, & Lachapelle, 2016; Pekrun & Bühner, 2014). Moreover, given that emotions are considered to be subjective experiences (Shuman & Scherer, 2014) that can be communicated (Pekrun & Linnenbrink-Garcia, 2014) and are composed of a *feeling* component, self-reports are well suited for this particular construct. A host of studies have found significant relations between self-reported emotions and performance in a variety of achievement contexts (e.g., Kavussanu, Dewar, & Boardley, 2014; Levine et al., 2011; Pekrun & Linnenbrink-Garcia, 2014). However, even the most prevalent and well-validated scales present challenges when measuring emotions within medical education.

For instance, the Positive Affect and Negative Affect Schedule (Watson, Clark & Tellegen, 1988; Watson & Clark, 1999) has been widely used within the broader emotion literature to measure affective states and discrete emotions. One limitation of this scale is that it does not fully represent the arousal dimension of emotions (i.e., activation and deactivation) for both positive and negative states. In addition, the discrete emotional states included in this scale were not selected for their relevance in achievement or educational settings, but rather are noted for their associations with mental health. As such, this scale excludes states that are likely to be activated in contexts in which performance outcomes and standards for success are more salient (e.g., pride). Furthermore, the discrete emotional states in this scale appear to represent broader blends of affective states that more directly represent states of cognition and alertness (e.g., *attentiveness, fatigue*) or traits and moods (e.g., *shyness, loneliness*) rather than emotional states, which are shorter-lived and have a stronger object focus.

Within the education literature, the Achievement Emotion Questionnaire (AEQ; Pekrun, Goetz, & Perry, 2005, 2011) has played a critical role in advancing understanding of the impact of emotions in academic settings (e.g., Goetz, Pekrun, Hall, & Haag, 2006; Pekrun et al., 2007, 2010, 2011). However, several features of this measure reduce its relevance for medical education. First, the questionnaire is tailored toward activities in traditional classroom learning environments and includes items containing statements about "studying," "class," "reading material," and "preparing for exams," that may not be relevant for problem-based medical learning environments, such as simulation and on-site clinical training. The AEQ makes a distinction between learning and performance activities, whereas medical education tasks that involve training-in-action may represent a blend between these facets of achievement. Similarly, the clauses attached to each emotion statement (i.e., antecedents, object-focus, actions) may not represent the way an emotion is experienced within medical culture, which may lead participants to agree with some aspects of the statement while disagreeing with others.[26] In addition, assumptions about the onset of specific emotions may operate differently in medical education. For instance, the AEQ outlines hope as occurring before or during an activity but not after it, whereas relief is expected after an activity but not during it. However, these assumptions need to be tested for medical tasks given that emotions may be activated at multiple times in anticipation of future events. Finally, although this scale includes emotions relevant for achievement tasks, it does not incorporate other states likely to be activated in settings involving social interactions (e.g., compassion, gratitude) or knowledge-based activities (e.g., confusion, curiosity).

With respect to measures developed or used specifically for medical education, existing affect-related scales mainly assess traits that relate to personality, interpersonal skills, and broader affective states, such as emotional intelligence (e.g., Chew, Hassan, & Zain, 2013), burnout (e.g., Dyrbye et al., 2014), and empathetic qualities (e.g., Dehning, et al., 2013) rather than emotional *states*. The most commonly used measure of state emotion in medical education is the State Trait Anxiety Inventory (STAI, Marteau et al., 1992; Spielberger, 1989; Spielberger, Gorsuch, Lushene, Vagg, & Jacobs, 1983), which has been used across a number of studies to examine trainees' task anxiety (e.g., Arora et al., 2010; LeBlanc et al., 2005, LeBlanc &

[26] For instance, the statement: "I felt so embarrassed I wanted to run and hide" reflects a motivational component that may not represent how medical trainees would respond to feelings of embarrassment in a culture that values confidence. Another statement reads: "I was happy that I could cope with the task." However, there may be other reasons or causes for happiness that would lead an individual to disagree with this statement.

Bandeira, 2007). However, there is also a need to consider a diverse array of emotional states (both positive and negative) as these may differentially affect learning and performance (Shuman & Scherer, 2014).

Across these bodies of literature, self-report measures of emotions differ in whether they measure states or traits; discrete emotions (e.g., anxiety) or broader emotion dimensions (e.g., negative states); emotion intensity or frequency; single or multiple emotions; Likert or visual analogue scales ratings, adjectives to assess an emotion feeling or a series of statements to assess the underlying emotion components. A review of the advantages and disadvantages of these design choices is beyond the scope of this paper (see Pekrun & Bühner, 2014); however, in the following section we explain our choices in the development of a new scale.

A New Emotions Scale for Medical Education

To develop the Medical Education Emotion Scale (MEES, Appendix G), we first considered which emotions should be included. We selected emotion taxonomies based on their relevance to medicine education. This included an array of emotions from taxonomies of *achievement emotions* (e.g., anxiety, enjoyment, Pekrun et al., 2011), *social emotions* (e.g., compassion, gratitude, Fisher & van Kleef, 2010; Weiner, 2007), *epistemic/knowledge emotions* (e.g., confusion, curiosity, Pekrun & Stephens, 2012), and *basic emotions* (e.g., fear, anger, sadness, Ekman, 1992). We argue that these emotions are likely to be activated in authentic medical education environments given the inherent performance expectations, social dynamics, high-stakes activities, and knowledge-based nature of medicine. This approach is consistent with the notion that emotions are considered domain-specific and may vary according to the educational environment (D'Mello, 2013; Goetz, Frenzel, Pekrun, Hall, & Lüdtke, 2007; Goetz, Pekrun, Hall, & Haag, 2006) and socio-cultural climate (Turner & Trucano, 2014).

The scale was designed to measure state (momentary) rather than trait (habitual) emotions given that we consider emotions to be continuously unfolding in real time and there is a need to measure experiences that occur during authentic learning, rather than focusing on trait-based aptitudes or tendencies, which may be more closely tied to stereotypes than moment-to-moment feelings (Goetz et al., 2013, Mauss & Robinson, 2009). As such, our goal was to create an efficient and practical scale that could be administered at various time points for authentic achievement activities without causing undue burden or interference with the task at hand. Although this precluded us from including an exhaustive list of emotional states, we contend that the collection of emotions in the MEES offers a more representative and meaningful range of emotions for medical education than existing alternatives.

Scale instructions were contextualized to both the activity at hand and specific time points to ensure a consistent frame of reference for reporting emotions and to limit the likelihood that participants reported more general moods, traits, or levels of fatigue. To limit scale length, we used single-item adjectives to measure each discrete emotion. This approach is consistent with previous research that has measured emotional states over time (e.g., Goetz et al., 2014; Nett, Goetz, & Hall, 2011). Furthermore, in their analyses of short-and long-form questionnaires of affective constructs, Gogol et al. (2014) found that, consistent with earlier research (e.g., Gardner, Cummings, Dunham, & Pierce, 1998), the psychometric properties of single-item scales are sufficiently sound compared to their multi-item counterparts and present a viable alternative when questionnaire length is a concern.[27] We chose adjectives (e.g., frustrated) instead of context-dependent statements (e.g., I felt frustrated when I encountered a complication in the procedure) as we did not want to make assumptions about which stimuli were linked to specific emotions. Instead, we elicited this type of information through interviews to understand antecedents and object-focus of emotions. This allowed us to measure the intensity of an emotion rather than the extent to which the participant agreed it occurred in concert with a specific event.

Given our choice to use adjectives and Likert ratings of intensity, the scale most directly measures the feeling component of emotions. This refers to the subjective experience and awareness of an emotional state that involves recognizing a change in one or multiple components that compose an emotion, such as feeling motivated to approach a stimulus, cognitively engaged, or physiologically activated (Ellsworth, 2013; Moors, 2014; Moors, Ellsworth, Scherer, & Frijda, 2013). Previous research has shown that across individuals, there is significant convergence in the underlying meaning of these emotion feelings. Specifically, researchers have found evidence for consistency in associations between emotion words and component profiles (cognitive, motivational, physiological features) across cultures and languages, suggesting that a consistent frame-of-reference exists for discrete emotion words in everyday life (Fontaine et al., 2013; Shuman & Scherer, 2014; Scherer, Shuman, Fontaine, & Soriano, 2013) and achievement settings (Loderer et al., under review).

[27] In future research, we aim to develop and test a long-form of the MEES to test the robustness of this finding using multiple adjectives for each discrete emotion. In this study, we focus on the short-form of the MEES given concerns that a longer questionnaire could interfere with task operation or performance.

Although some researchers contend that the broader dimensions of valence (positive versus negative) and arousal (activating versus deactivation) are more robust than distinctions between discrete emotions (Posner, Russell, & Peterson, 2005), we adopt the stance that discrete states and dimensional features are not incompatible (Izard, 2007; Maus & Robinson, 2009; Scherer, 2005; Shuman & Scherer, 2014). Thus, the MEES allows us to examine both by presenting discrete emotions that can be classified into categories according to valence and arousal. Assessing the validity of this scale involves gathering evidence to test convergence between scale responses, related psychological constructs, and other measures of emotions. In the following section, we discuss how we set out to accomplish this in the present research.

Aims of the Present Research

Given limitations in existing self-report scales, there was a need to develop a measure of emotions that could be administered for authentic learning and achievement tasks within medical education. Thus, we developed the Medical Education Emotion Scale (MEES), which includes emotion taxonomies specifically selected for their relevance to medicine education. The purpose of this research was to: (1) examine the initial validity of a self-report scale (the Medical Education Emotion Scale, MEES); (2) test the applicability of control-value theory within medical education; and, (3) identify areas for further inquiry in emotion theory and measurement that could aid future scale development and research in medical education.

As St-Onge and Young (2015) argue: "there may not be a holy grail of validity theory or validation framework for assessment in health professions education." (p. 548). Instead, they suggest that the validation process and the types of evidence gathered may differ depending on the context and construct being assessed.[28] In the present study, our goal was to develop a scale that could be used in real-world medical training beyond the classroom, including simulation and clinical settings. Therefore, we chose to focus on the following aspects of validity (see McIntire & Miller, 2007): *criterion validity* (i.e., measure relates to—concurrently or predictively—theoretically related constructs or behaviors), *convergent validity* (i.e., measure relates to other measure of the same or similar construct), *face validity* (i.e., measure is viewed by participants to represent intended construct), and *content validity* (i.e., measure assesses adequate range/facets of a construct). We also examined *cognitive validity* (Karabenick et al., 2007), which is often a neglected aspect of validity, but particularly relevant for self-report scales as it examines whether

[28] For discussions of validition frameworks for measures of *assessment* in medicine, see Cook et al., 2014, 2015

participants' interpretation of scale items and response behaviors are coherent and consistent with researcher assumptions (i.e., whether it aligns with the assumed cognitive response process), which can be considered a component of construct validity.[29]

More specifically, in Study 1 our aim was to address the following research questions using retrospective and cognitive interviews:

1. Do scale responses align with interview responses? (convergent validity)
2. Does the scale represent an adequate range of emotions? (face, content validity)
3. How do participants interpret and respond to scale items? (cognitive, construct validity)
4. What distal antecedents are linked to emotions? (construct validity, scale development)

In study 2, our aim was to address the following research questions by administering the MEES across three authentic medical education environments:

5. Do emotions measured using the scale relate to theoretically related constructs (appraisals, performance, post-task self-efficacy)? (criterion-related validity)
6. Which distal antecedents are commonly reported? (construct validity, scale development)
7. Are the underlying components and profile structures of the scale consistent with theory or do they identify areas requiring further inquiry? (construct validity, scale development)

In terms of hypotheses, we expected emotions reported via the MEES to align with emotions reported via interviews (question 1). We expected participants to perceive the scale to be representative of their emotional experiences (question 2). In terms of relations to psychological constructs (question 5), we expected control and value to positively relate to positive emotions, and control to negatively relate to negative emotions. We also expected positive emotions to positively relate to performance and self-efficacy, whereas negative emotions would negatively relate to performance and self-efficacy (see Pekrun, 2006). We did not formulate specific hypotheses for the remaining questions (questions 4, 6, 7) as these were more exploratory in nature and analyses were conducted primarily to aid further scale development and research.

Study 1

The goal of study one was to use interviews to examine convergent, face, and cognitive validity and to identify antecedents that could aid further scale development.

[29] These validation categories are not mutually exclusive, but rather provide a useful way of gathering and organizing evidence to test and establish construct validity (Camara & Lane, 2006; Karabenick et al., 2007).

Method

Participants

Fifteen medical trainees[30] (n = 10 female) participated in the computer-based learning environment, BioWorld. Of these, 26.7% identified as Caucasian, 26.7% as Asian, 20.0% as Arab/Middle Eastern, and 26.7% as other/mixed ethnicity. The mean age was 24.33 (SD = 3.44) and the mean year of medical school was 3.73 (SD = 1.33).

Learning environment.[31] Medical trainees were presented with two medical cases (diabetes and pheochromocytoma) counterbalanced within BioWorld (Lajoie, 2009), a computer-based learning environment designed to support diagnostic reasoning skills by solving patient cases (see Figure 9). Learners could view patient history, order simulated lab tests, and use a digital library tool. These features provided medical information and scaffolding to help learners determine which diagnosis the evidence best supports. After submitting their final diagnosis, learners received feedback on the selection of evidence based on an expert solution.

Measures

Demographics. Demographics were collected including age, gender, ethnicity, and level of medical training/practice (e.g., year of education).

Emotional states. The Medical Education Emotion Scale (Appendix G) was administered to measure the intensity of 22 emotions using single-item adjectives for discrete emotions (e.g., anxious, proud, relieved, frustrated) belonging to four categories of emotion taxonomies: basic, epistemic, social, and achievement. The following four sub-scales were created based on valence (pleasant/unpleasant) and activation (level of physiological arousal) of emotions: (1) positive activating (e.g., enjoyment, pride, curiosity), (2) positive deactivating (e.g., relaxation, relief), (3) negative activating (e.g., anxiety, frustration, confusion), and (4) negative deactivating emotions (e.g., hopelessness, disappointment). Neutral and surprise were included as non-valenced states. A five-point Likert scale was used to measure intensity ranging from 1 (*not at all*) to 5 (*very strongly*).

Procedure

Participant consent was secured prior to data collection. Participants were administered the MEES immediately before engaging in the diagnostic reasoning task within the BioWorld

[30] Thirteen medical students and 2 med prep students.
[31] Data in BioWorld as collected as part of a larger research program in which participants completed two cases. Data presented here includes only second case responses as this facilitated a clearer reference points for interviews.

environment, every 10 minutes during the task (approximately 2-3 times), and immediately after the task was completed (after receiving feedback). The mean score of emotions reported during the learning session was averaged for analyses, as this was considered to best represent the cummalative emotions experienced during the task and would facilitate analyses of alignment between scale and interview responses. Afterward, a one-on-one interview (approximately 15 minutes) was conducted with each participant. Retrospective interviewing (Ericsson & Simon, 1993) and cognitive interviewing techniques (Muis et al., 2013) were employed following the learning session (see Table 14 for sample questions, Appendix H for complete list and instructions).

The purpose of the retrospective interview was to ask participants to recall the state emotions they were experiencing at various points in the learning session (before, during, after) as a method of establishing convergent validity (assessing alignment between interview and MEES responses). This also provided a better understanding of the antecedents and object-focus of these emotions as a guide for future scale development and criterion-related validity. The purpose of the cognitive interview was to ask participants to explain their interpretation and response to scale items[32] as a method of establishing construct, content, and face validity (assessing whether the MEES captured a representative range of emotions, whether it was interpreted in a manner that aligned with researchers' assumptions). The interview questions, protocol, and initial coding scheme were first piloted with a separate sample of medical trainees ($N = 5$) who participated in the BioWorld environment following the protocol described above. Based on this pilot data, some questions were reworded to improve clarity or eliminated if deemed too challenging. Interviews were conducted by two interviewers who were trained to use the original and revised protocol. To limit response biases and build rapport (e.g., social desirability, ad-hoc explanations), participants were reminded that there were no right or wrong answers and that our goal was to understand the emotions they experienced and their perceptions of the scale. We also encouraged participants to describe the emotions they recalled experiencing rather than to try to keep interview responses consistent with scale responses.

[32] Cognitive interview questions were aimed initially at the scale level (i.e., how participants typically responded, what they found challenging or unclear overall), rather than at the item level (cf. Karebenick et al., 2007). However, follow-up prompts ("Was there an item in particular? Were there any exceptions?"), as well as participant responses also elicited item-level responses (e.g., the item *neutral* being perceived as particularly difficult to interpret).

Interviews were transcribed and coded for the type of emotion reported, associated antecedents, and expressions[33] (see Table 15 for coding scheme). Transcripts were also coded for participants' perception of scale clarity, relevancy of emotions, and interpretation of items. Two raters independently coded a subset (14%) of the interview transcripts to establish inter-rater reliability for emotion codes (before, during, after emotions) used to examine alignment between the interview and the MEES. Segments and codes were compared for instances of agreement/disagreement. High inter-rater reliability was established ($\varkappa = .86$). Discrepancies were resolved through discussion.

Results

Group-level Alignment (Convergent Validity)

Retrospective interviewing results demonstrated alignment between MEES and interview responses at the group-level (across participants) and individual-level (within participants) for the task. A group-level comparison between the most intensely reported emotions on the MEES (Table 16) and the most commonly reported emotions in retrospective interview responses (Table 17) demonstrates that across participants, interview and questionnaire responses were 100% aligned for each phase of the diagnostic reasoning task. A match was identified when the most commonly reported emotion experienced in the interviews across participants (based on group-level percentage) aligned with at least one of the highest ranked positive and negative emotions in the MEES across participants (based on group-level mean).[34]

For example, in the interview the majority of participants reported feeling curious *before* the task (56%); this was also the top rated positive emotion on the MEES before the task (M = 3.9, SD = 0.6). In terms of negative emotions at this phase, in the interview participants most commonly reported feeling anxious/stressed (29%[35]); anxiety was also the top rated negative emotion on the MEES before the task (M = 2.2, SD = 0.9). For the *during* phase, the majority of participants reported in the interview that they recalled feeling happiness/enjoyment (73%); enjoyment was also the top rated emotion during the task (M = 3.7, SD = 0.9). In terms of negative emotions at this phase, the majority of participants recalled feeling confused/frustrated (33%); confusion was also the top rated negative emotion on the MEES during the task (M = 2.4,

---

[33] Expression codes were not analyzed for this study.
[34] In the case of two or more emotions with the same mean score
[35] In the interview, some participants did not report experiencing negative emotions for certain phases of the task—or only to a small degree.

SD = 1.2). Finally, for the *after* phase, most participants recalled feeling happiness/gladness (66%); enjoyment was also the top rated positive emotion on the MEES after the task (*M* = 3.3, *SD* = 0.9). In terms of negative emotions after the task, the majority of participants in the interview recalled feeling confused/frustrated 27%; confused was also the top rated emotion on the MEES after the task (*M* = 1.9, *SD* = 0.8). Table 17 illustrates sample quotations and object-focus of reported emotions.

Individual-level Alignment (Convergent Validity)

A comparison between individual interview responses and corresponding highest rated emotions on the MEES demonstrates that there was 86% alignment between the interview responses and the MEES for emotions reported *before*, 80% alignment for emotions reported *during*, and 73% alignment for emotions reported *after* the diagnostic reasoning task. Table 18 illustrates examples of alignment/misalignment for a sample of participants. A match was identified for each individual when the most intense emotion(s) in the interview aligned with at least one of the most intense emotions reported on the MEES for the corresponding phase of the task. Typically, examples of misalignment occurred when several emotions were rated at the same level of intensity (e.g., 3 on the Likert scale) and there was no clear top ranked emotion or when the ratings of emotional intensity were generally low across multiple emotions, which suggests that more intense unique emotional states may be easier to recall due to their salience. In other cases of misalignment, participants combined multiple emotions that may have referred to a similar affective state (e.g., reported *sad* and *frustrated* in interview versus *disappointed* on MEES) or referred to broader affective states that may have encompassed the top rated emotion (e.g., reported *happy* in interview versus *proud* on questionnaire). Additional follow-up prompts for further contextual information or explicit comparisons to questionnaire ratings during the interview may resolve these discrepancies.

Cognitive Interviews (Face and Construct Validity)

Content analyses from cognitive interviews are reported in Table 19. Findings revealed that participants found the MEES captured the range and variability of emotions experienced during the diagnostic reasoning task at an appropriate frequency. Participants found the scale straightfoward and clear to understand. However, it was suggested that *excitement* and *embarrassment* be added to the questionnaire (although similar affective states, such as curiosity/enjoyment/happiness and disappointment/shame were included).

In terms of emotions that were considered irrelevant, participants identified *disgust, compassion, anger,* and *shame* as less relevant for this context. *Compassion* was seen as less relevant within a computer-based environment;[36] however, participants suggested that they would likely feel greater compassion in a clinical setting involving face-to-face interactions with patients. Some participants considered *anger* and *shame* to be too intense. Disgust was difficult for participants to pair with an object-focus. In terms of emotions that were considered challenging to interpret, *neutral* was commonly cited by participants. The main issue appeared to be that participants found it difficult to conceptualize different intensities of a neutral state (e.g., "intensely neutral").

Regarding interpretation of the scale, most participants reported that they rated their emotions in terms of how intensely or strongly an emotion was felt rather than the frequency or duration, which is consistent with the intended design. However, interpretation of the mid-way point of the scale was more varied. For example, some participants considered the mid-way point on the Likert scale ("moderate") to reflect a baseline level for a given emotion, while others felt that it reflected an emotion experienced but not intensely, and yet others still selected it when they were uncertain. In some cases, a different response approach was used depending on the type of emotion. For example, a participant may rate their trait-like feeling for certain emotions (e.g., typically happy disposition translated into consistent high ratings of happiness), whereas for other emotions, the same participant may rate their state-like feeling (e.g., increase in anxiety compared to the beginning). This suggests that certain emotions may be more susceptible to trait-based dispositions.

Antecedents (Construct Validity and Scale Development)

Content analyses of retrospective interviews also revealed several key antecedents of emotion activation including: the nature of the task and learning environment (e.g., complexity or novelty of case), information related to the case (e.g., lab test results), knowledge and skills (e.g., recognition of symptoms, problem-solving strategies), as well as performance (e.g., past, current, or anticipated performance).

In terms of links between antecedents and emotions, task/environment features elicited positive emotions (e.g., curious, excited, relaxed, relieved) when participants felt the learning

[36] Despite this response we retained this emotion in the scale given that it is likely to be activated in real-world clinical settings or team-based environments.

environment was supportive or when they found the task to be meaningful and not overly complex. For example, as one participant stated: “I was happy to be here and I was not feeling that pressured.” However, participants noted feeling negative emotions (e.g., confused, anxious) when they felt the task was too complex or couldn’t locate the information they were seeking. As one participant noted: “I remember being a bit confused because it was a more complex case.”

Knowledge was linked to positive emotions (e.g., hopeful, proud) when participants felt they understood or had sufficient knowledge to solve the case. As one participant claimed: “I felt a bit proud because I knew what I had to look for.” On the other hand, participants reported negative emotions (e.g., frustration, confusion, guilt) if they felt they lacked sufficient knowledge to solve the case. As one participant stated: “I was frustrated that I didn’t know the indications for that.”

Simulated lab test results were typically associated with positive emotions (e.g., excited, proud, hopeful), when the evidence supported an existing hypothesis or helped to make a decision. As one participated stated: “I was happy because it gave me a hypothesis.” On the other hand, participants typically responded with negative emotions (e.g., confused, afraid, disappointed) when they were unclear about how to interpret information (e.g., unanticipated results) or when the information challenged their current hypothesis without providing a definitive direction. For instance, as one participant explained: “[I was] disappointed because I thought I was getting to something that was more concrete.”

In terms of performance, participants recalled feeling positive emotions (e.g., proud, relaxed, excited) when they were satisfied with their past performance, when their current performance provided reassurance that they were progressing efficiently, or when they anticipated a positive outcome (e.g., correct diagnosis). As one participant commented: “I was happy and kind of proud that I came up with the right diagnosis.” In contrast, participants described feeling negative emotions (e.g., anxious, confused) when they were not satisfied with their previous performance, when they were unsure about their current performance (e.g., inefficiency, choice of lab tests), or anticipated negative outcomes (e.g., incorrect diagnosis). For example, one participant commented: “I was also kind of afraid that I was completely off.” In some cases, anticipated future performance moved beyond the case and learning environment outcomes (e.g., concern about whether they would be able to effectively and efficiently make accurate diagnoses in future real-world clinical settings). Alignment to expert analysis/feedback after submitting diagnostic solution (e.g., correct diagnoses, similar use of evidence to justify decision) appeared to play a key role in forming these performance standards.

Discussion

Retrospective interview responses revealed that there was strong alignment between the emotions reported though concurrent questionnaires and retrospective interviews. These results suggest that the MEES can be used to accurately capture the range and intensity of emotions experienced by medical trainees in authentic medical contexts. The results from the retrospective interview also revealed important emotion antecedents. Specifically, emotions were triggered in response to the task/environment, the simulated laboratory test results, perceptions of knowledge, and performance (past, current, or anticipated). These antecedents as well as the object-focus of emotions described in participants' responses can guide the wording and development of longer-form scales using more traditional statement items rather than adjectives to assess emotions in medical education. In addition, these findings can help to further refine control-value theory by delineating specific distal antecedents that may be particularly relevant for medical education.

Cognitive interview responses revealed that the emotions included in the MEES captured the range of emotions participants experienced in the medical environment, which helps to establish face validity. Further corroborating these findings, descriptive statistics from the MEES demonstrated that the most intensely reported discrete emotions represented a range of emotions from each taxonomy (i.e., basic, social, epistemic, achievement). In addition, the commonly reported behavior of selecting response choices based on the *intensity* of the emotion was consistent with the intended design of the scale, which provides evidence supporting construct validity. Participants generally felt the scale was meaningful, clear, and straightforward to complete. However, cognitive interview responses also provided some insights that could guide further refinement of the scale and improve clarity. For instance, given variations in participants' interpretations and response behaviors toward the mid-way point of the scale in particular, adding more descriptive anchors or an example in the instructions, may help to improve consistency. In addition, adding emotions such as *excitement* and *embarrassment*, as well as removing items, such as *disgust* and *neutral* could increase the utility and relevance of the scale.

Although we did not make these changes for study 2 given that data was collected concurrently in other medical environments, these insights can help to guide future scale design and measurement. Furthermore, the interview and analytic techniques used in this study can be employed in future research that aims to examine the validity of various emotion measures. Researchers could also consider examining a trait-based measure of emotions, in addition to state emotions, to determine whether certain emotions are more likely to be influenced by dispositions or habitual emotional experiences. Finally, examining additional validity evidence, such as discriminant validity (whether measure is distinct from theoretically unrelated constructs) and criterion-related validity (whether measure relates to theoretically related constructs) would help to further bolster the evidence presented here. In Study 2, we address this latter form of validity, as well as explore the underlying structure of emotions measured using the MEES.

Study 2

Study 2 expanded on the findings of Study 1 by examining the criterion-related validity of the scale across three environments and, in doing so, also tested the generalizability and utility of control-value theory in medical education. A secondary goal was to conduct exploratory analyses to examine the underlying component structure and profiles for future inquiry and scale development. For the purposes of this study, our goal was to pool data from authentic environments rather than examine differences between environments to test the generalizability of control-value theory and examine the nature and structure of emotions using a larger sample.

Method

Participants

Sixty medical trainees volunteered to participate in this research. Participants were recruited from a public university and teaching hospital in North America. Sixteen medical residents and final-year students (n = 3 female) conducted a surgical procedure within an operating room under supervision. The mean age was 30.31 (SD = 5.51); mean year of residency was 2.19 (SD = 2.20). Twenty-four second-year medical students (n = 9 female) solved virtual patient cases within a computer-based learning environment. The mean age was 24.46 (SD = 3.54). Twenty fourth-year medical students (n = 14 female) participated in emergency care role-play simulations. The mean age was 27.53 (SD = 4.06).[37]

Learning Environments

We selected three environments that represented distinct authentic learning environments, providing a range of medical education activities. Each environment represented an authentic medical learning environment given that they each involved problem-based

[37] Ethnicity data is not presented here as it was not collected within the deteriorating patient environment.

tasks/activities, context-sensitive measures of skills/knowledge and took place in either real-world clinical practice or a simulation counterpart. All provided opportunities to improve medical knowledge/skills at the same time as providing an opportunity to demonstrate competence. In this way the activities in these environments can be considered both learning tasks (improve skills/knowledge) and achievement tasks (involves a success or failure outcome). Differences and similarities in various features of the environments (e.g., instructional design, learning goals, social dynamics, fidelity, feedback) are summarized in Table 20.

Surgical environment. Medical trainees conducted an open saphenous vein harvest procedure during cardiac surgery under supervision of medical staff (Duffy et al., 2015b). This is a technical procedure in which a vein is harvested from the leg and later used as a graft during coronary artery bypass surgery to circumvent blocked arteries and restore normal blood flow in the heart (see Figure 10). This training provided medical trainees with an opportunity to improve technical skills and learn about surgical procedures in a naturalistic setting.

BioWorld environment.[38] Medical trainees were presented with one of three medical cases (diabetes, human immunodeficiency virus, pheochromocytoma) within BioWorld (Lajoie, 2009), a computer-based learning environment designed to support diagnostic reasoning skills by solving patient cases (see Figure 9). Learners could view patient history, order simulated lab tests, and use a digital library tool. These features provided medical information and scaffolding to help learners determine which diagnosis the evidence best supports. After submitting their final diagnosis, learners received feedback on the selection of evidence based on an expert solution.

Deteriorating patient environment.[39] Medical trainees engaged in a role-play simulation to learn to diagnose and stabilize patients with deteriorating medical conditions using a medical algorithm (e.g., airway, breathing, circulation; Wiseman & Snell, 2008). Within groups, students took turns in the role of physician to stabilize the patient before cardio respiratory arrest occurred. The medical tutor played the role of patient and nurse. There were five medical cases revolving around three conditions: gastro-intestinal bleed, anaphylaxis, and opioid toxicity. Debriefing after cases provided individualized feedback.

---

[38] Data collected in this environment was part of a larger research program in which participants typically completed two cases. To limit confounds in our analyses across environments, we analyzed data from the first case completed.
[39] Data collected in this environment was part of a larger research program that involved participating in simulations over two days. We analyzed data from the second day as this was closer in time to the performance measure.

Measures

Demographics. Demographics were collected including age, gender, and level of medical training/practice (e.g., year of education).

Emotional states. The Medical Education Emotion Scale (see Figure 11, Appendix G) was administered to measure the intensity of 20 emotions[40] using single-item adjectives for discrete emotions (e.g., anxious, proud, relieved, frustrated) belonging to four categories of emotion taxonomies: basic, epistemic, social, and achievement. The following four sub-scales were created based on valence (pleasant/unpleasant) and activation (level of physiological arousal) of emotions: (1) positive activating (e.g., enjoyment, pride, curiosity), (2) positive deactivating (e.g., relaxation, relief), (3) negative activating (e.g., anxiety, frustration, confusion), and (4) negative deactivating emotions (e.g., hopelessness, disappointment). Surprise was included as a non-valenced activating state. A five-point Likert scale was used to measure intensity ranging from 1 (*not at all*) to 5 (*very strongly*). The internal reliability levels for each sub-scale (collapsed over time points) were as follows: positive activating (α = .91), positive deactivating (α = .69), negative activating (α = .92), negative deactivating emotions (α = .88).

Appraisals. Control (Pekrun, 2006) was assessed using a single item to measure cognitive appraisals of control toward performance. Value (Wigfield & Eccles, 2000) was assessed using a single item to measure perceived value of the task.[41] For both items, a 5-point Likert scale was used ranging from 1 (*strongly disagree*) to 5 (*strongly agree*).

Self-efficacy. Post-task self-efficacy was assessed using a single item adapted from Bandura's (2006) scale to measure participants' confidence that they could effectively conduct the procedure/do well on a similar task in the future.[42] A Likert scale was used to measure confidence ranging from 0 (*no confidence*) to 100 (*complete confidence*) with 10-point intervals.

Performance. Performance was assessed using context-specific measures for each environment. In the surgical environment, the Objective Structured Assessment of Technical Skill (OSATS) global rating scale (Martin et al., 1997; Reznick et al., 1997, Appendix I) was used by two expert raters to assess multiple facets of *technical* skills, including efficiency,

---

[40] Based on participant feedback from Study 1, we did not include *neutral* or *disgust* in analyses

[41] A composite score was used for value (composed of interest, importance, utility) in the surgical environment.

[42] Although self-efficacy can be considered a facet of control (Pekrun et al., 2007), it has also been studied as a key factor in motivation (Pajares, 1996). In this study, it can be considered an aspect of control that concentrates on the *confidence* one feels in their competence for future tasks. Another way to interpret this variable is as a proxy for post-task control, and thus, as a means of examining how emotions influence appraisals toward future events.

instrument handling, and knowledge of procedure (mean score of raters' total score was used in analyses, intra-class correlation = .95). In the BioWorld environment, *diagnostic reasoning skills* were assessed based on percent agreement between participants' prioritized evidence and an expert's diagnostic solution (see Appendix J).[43] Diagnostic solutions for each case were based on subject-matter consultations and task analyses (Gauthier & Lajoie, 2014). Performance scores were obtained via log-files. For the deteriorating patient simulation, performance was assessed through a multiple-choice post-test that assessed participants' *procedural knowledge* to stabilize critically deteriorating patients (Wiseman & Labelle, 2015; Appendix K). These tests were developed by a team of subject-measure experts. These measures were selected given that they were best suited to assess the skills that each learning environment was designed to foster: namely, technical skills, diagnostic reasoning skills, and procedural knowledge.

Prior knowledge. Prior knowledge/experience was assessed in each environment. In the surgical environment, participants reported the number of times they had previously conducted the vein harvest procedure. In the BioWorld environment, participants reported their prior knowledge of the case topic using a Likert scale ranging from 1 (*no knowledge at all*) to 7 (*a great deal of knowledge*). In the deteriorating patient simulation environment, participants completed a pre-test, which measured procedural knowledge.

Antecedents. Antecedents of emotions (i.e., factors contributing to the onset of emotions) were measured in the BioWorld environment using a checklist inventory composed of the following categories: case complexity, prior knowledge, problem-solving strategies, anticipated/current performance, information acquired during the task (e.g., lab test results), features of the learning environment (e.g., design of computer-based learning environment), expert feedback, and other. Participants were asked to check items that had contributed to the emotions they experienced before, during, or after the task. These factors can be considered distal antecedents of emotions, as they are likely to impact the immediate task appraisals (e.g., control, value), in turn impacting emotions (Pekrun & Perry, 2014). The list was generated based on our analyses of retrospective interviews in BioWorld with a separate participant sample (Study 1).

Procedure

Figure 12 provides an illustration of the research protocol. Participant consent was

[43] Performance data was standardized for each of the three cases in BioWorld.

secured prior to data collection. Task instructions, demographics, prior knowledge, and appraisal scales were administered immediately before the task.[44] Participants completed emotional states scale in reference to three time points (before, during, after the task).[45] Participants' self-efficacy was measured after task completion. Antecedents for emotions were analyzed after task completion in the BioWorld environment. Performance scores were analyzed after task completion. Tasks took varying lengths of time to complete (approximately one hour) depending on the environment; however, performance scores were standardized first for each case in BioWorld and then for each of the three environments in order to make comparisons for correlation analyses and to pool data into a larger sample to test regression models. Prior knowledge/experience scores were also standardized for each of the three environments.

Results

Correlation Analyses[46]

Bivariate correlation analyses were conducted within each environment to determine whether expected relations occurred within each environment. The predictor variables that demonstrated the strongest and most consistent correlations to criterion variables were selected for inclusion in subsequent regression analyses (Greene et al., 2013, 2015). Descriptive statistics for cognitive appraisals and emotions (based on taxonomy and valence groupings) are provided in Table 21 (statistically significant differences between environments are indicated).

Surgical environment. Within the surgical environment, results revealed that control was negatively related to negative activating emotions ($r = -.46, p < .05$) and negative deactivating emotions ($r = -.42, p = .05$). Control was also positively related to performance ($r = .46, p < .05$). In terms of relations between emotions and task outcomes, results revealed that negative activating emotions were negatively related to post-task self-efficacy ($r = -.47, p < .05$). See Table 22 for correlation matrix.

BioWorld environment. Within BioWorld, results revealed that control was positively related to positive activating emotions ($r = .44, p < .05$) and positive deactivating emotions ($r =$

[44] With the exception of prior knowledge in BioWorld. In this environment, prior knowledge was reported afterwards (i.e., self-assessed post-hoc) to ensure that participants were not biased in their diagnosis of the case.
[45] *During* questionnaires were administered after the task to prevent interference with the task. In Study 1 we found that retrospective accounts of emotions were highly aligned to emotions measured concurrently for activities lasting approximately 1 hour.
[46] Although not presented here, correlation analyses were examined for emotions at each time point. Here we report analyses of averaged scores for each emotion sub-scale collapsed over time to reduce redundancies.

.47, $p < .05$) and negatively related to negative deactivating emotions ($r = -.33$, $p = .055$). Control was also positively related to performance ($r = .42$, $p < .05$). In terms of relations between emotions and task outcomes, positive activating emotions were positively related to performance ($r = .32$, $p = .067$) and post-task self-efficacy ($r = .71$, $p < .01$). Positive deactivating emotions were also positively related to post-task self-efficacy ($r = .67$, $p < .01$). Negative activating emotions were negatively related to performance ($r = -.41$, $p < .05$) and post-task self-efficacy ($r = -.44$, $p < .05$). Negative deactivating emotions were also negatively related to performance ($r = -.44$, $p < .05$) and post-task self-efficacy ($r = -.46$, $p < .05$). See Table 23 for correlation matrix.

Deteriorating patient environment. Within the deteriorating patient (DP) simulation, results revealed that control was positively related to positive deactivating emotions ($r = .54$, $p < .01$). Control was also positively related to performance ($r = .47$, $p < .05$) and self-efficacy ($r = .72$, $p < .01$). In terms of relations between emotions and task outcomes, positive activating emotions were positively related to post-task self-efficacy, ($r = .52$, $p < .05$), positive deactivating emotions were also positively related to post-task self-efficacy ($r = .69$, $p < .01$). Negative activating emotions were negatively related to performance ($r = -.63$, $p < .01$). See Table 24 for correlation matrix.

Partial correlations. Partial correlation analyses were conducted to control for prior knowledge/experience. Results revealed that among all variables selected for entry into the regression model (see below), the relationships to outcomes remained statistically significant ($p < .05$) after controlling for prior knowledge/experience.

Hierarchical Regression Analyses

Based on correlation analyses, in model 1 (performance) we included negative emotions experienced before and during the task (entered first) and the control appraisal (entered second) to determine whether these combination of variables significantly predicted performance. In the second model (self-efficacy), we selected positive emotions before and during the task (entered first) and the control appraisal (entered second) to determine whether these variables significantly predicted post-task self-efficacy. Hierarchical regression analysis for performance revealed that a model including negative activating/deactivating emotions and control significantly predicted performance, $F(5, 55) = 4.35$, $p < .01$, and accounted for 25% of the variance (Adjusted $R^2 = .25$). In particular, negative activating emotions during the task ($\beta = -.66$,

$p < .01$) and control ($\beta = .30$, $p < .05$) were significant predictors of performance. Adding the control appraisal into the model provided a statistically significant increase in the amount of variance in performance that could be predicted by the model $F(1, 54) = 5.40$, $p < .05$. Hierarchical analysis for self-efficacy revealed that a model including positive activating/deactivating emotions and control significantly predicted self-efficacy, $F(5, 53) = 6.94$, $p < .001$, and accounted for 34% of the variance (Adjusted $R^2 = .34$). In particular, positive activating emotions during the task ($\beta = .52$, $p < .01$) and control ($\beta = .32$, $p < .05$) were significant predictors of self-efficacy. Adding the control appraisal into the model provided a statistically significant increase in the amount of variance in post-task self-efficacy that could be predicted by the model, $F(1, 53) = 6.67$, $p < .05$ (Table 25).

Antecedents

Results from analyses of the antecedent checklist in BioWorld revealed that the majority of participants reported that the following factors contributed to the emotional states they experienced before, during, or after the task: *prior knowledge* (88.69%), *case complexity* (80.36%), *information gained during the case* (60.71%), *anticipated or current performance* (57.14%), *feedback* (57.14%), and *problem-solving strategies* (54.17%) were most commonly selected. *Features of the learning environment* were reported by nearly half the participants (45.83%). The *other* category was reported in only one instance (3.57%).

Exploratory Analyses

Given that results from Study 1 demonstrated initial validity of *discrete* emotions, we conducted cluster and principal components analyses to examine the solutions (i.e., how these emotions related to one another) when grouped into distinct profiles (cluster analysis) and when reducing the total list of emotions to their essential dimensions (principal component analysis).[47, 48] We chose to conduct exploratory analyses rather than restricting the factor structure to previously identified structures with confirmatory factor analyses. This provided an

[47] Principal component analysis is a data reduction technique. Although other dimensions may be relevant, this analyses detects the most essential components by selecting the minimum number of uncorrelated components (i.e, orthogonal to one another) that account for the maximum variance (Meyers, Gamst, & Guarino, 2013).

[48] Factor analysis and principal component analysis represent two different causal/directional models: in factor analysis the factor causes the items; in principal component analysis, the items give rise to the component (aggregate descriptor). Given that it is unresolved whether sub-systems cause a discrete emotion (e.g., physiological arousal causes feeling anxious, James & Lange, 1922) *or* a discrete emotion causes the activation of sub-systems (e.g., feeling anxious cause physiological arousal, Cannon, 1927), either analyses at this exploratory stage can be justified. Moreover, it has been argued that both analyses are likely to produce similar solutions (Meyers et al., 2013).

opportunity to determine whether unique dimensions emerged given that the MEES contained a wider range of emotions than previously examined among a relatively untested sample.

Cluster analysis. To examine whether profiles could be formed based on emotions, we conducted three cluster analyses using the *k*-means procedure with a 3-cluster solution[49] on observed emotions scores of the MEES.[50] The three-cluster solution demonstrated interpretable results (i.e., distinct, non-redundant, profiles). Cases for Cluster 1 were higher in negative emotions compared to other clusters (with the exception of boredom before and after, which was higher on Clusters 2 and 3; curiosity before, which was higher on Cluster 1; and surprirse before, during, and after, which was higher on Cluster 1). Cases in Cluster 2 were relatively lower across all emotions compared to the other clusters. Cases in Cluster 3 were relatively higher in positive emotions compared to other clusters (with the exception of curiosity before, which was higher on Cluster 1, as previously mentioned). There were an adequate number of cases in each cluster (18-47% of total sample). Univariate ANOVAs were examined to verify whether the groups formed by these clusters differed significantly on both negative and positive emotions. Results revealed a significant difference between groups for negative emotions, $F(2, 57) = 58.96$, $p < .001$ and positive emotions, $F(2, 56) = 35.05$, $p < .001$. Specifically, post-hoc analyses revealed that participants in Cluster 1 demonstrated significantly more intense negative emotions than participants in Clusters 2 and 3 ($p < .001$). Participants in Cluster 3 demonstrated significantly more intense positive emotions than participants in Clusters 1 and 2 ($p < .001$). These results are consistent with our interpretation of the cluster profiles.[51]

Principal component analysis.[52] To investigate the structure of the emotions, we conducted three principal component analyses with varimax rotation, using the observed emotion scores of the MEES[50] for each time point (before, during, after task). Factor extraction was based on scree-plot investigation and the total percentage variance explained (acceptable at a minimum of 50%) after rotation, rather than restricting the factor solution.

The first solution (emotions before) pointed toward a three-factor structure that explained

---

[49] An unsupervised hierarchical cluster analysis using Ward's method and euclidian distance was conducted to identify the appropriate number of clusters through visual inspection of the dendrogram. 3 distinct clusters emerged.
[50] Excluded *disgust* and *neutral* as these were considered potentially difficult to interpret based on Study 1 results.
[51] These ANOVA results should be regarded as descriptive (verifcation or interpretation of cluster profiles) rather than tests of null hypotheses given that *k*-means clustering is intended to maximize differences (Meyers et al., 2013).
[52] Although some researchers have used principal component analysis prior to cluster analyses, there is evidence that in some cases this does not improve the cluster quality (see Yeung, Yee, & Ruzzo, 2000). Moreover, we consider these analyses to provide complementary exploratory approaches (i.e., case-oriented and variable/item-oriented).

50.74% of the total variance. After varimax rotation, the first factor accounted for 22.55% of the variance. Based on the six items that had their highest loadings on this factor (disappointment, hopelessness, sadness, anger, frustration, shame), it can be interpreted as representing a negative valence dimension. The second factor explained 16.88% of the variance. Based on the six items that had their highest loadings on this factor (pride, happiness, enjoyment, hopefulness, relief, relaxation), it can be interpreted as representing a positive valence dimension. The third factor explained 11.31% of the variance. Based on the three items that had their highest loadings on this factor (confusion, curiosity, surprise) it can be interpreted as representing an epistemic dimension, as exclusively epistemic emotions loaded the highest onto this factor.

The second solution (emotions during) pointed toward a four-factor structure that explained 58.42% of the total variance. After varimax rotation, the first factor accounted for 24.13% of the variance. Based on the six items that had their highest loadings on this factor (disappointment, hopelessness, sadness, anger, frustration, shame), it can be interpreted as representing a negative valence dimension. The second factor explained 13.40% of the variance. Based on the two items that had their highest loadings on this factor (anxiety, fear), it can be interpreted as representing a negative *activating* valence dimension. It should be noted, however, that this factor had several complex loadings and may not represent a unique factor. The third factor explained 12.03% of the variance. Based on the five items that had their highest loadings on this factor (gratitude, relief, pride, hopefulness, happiness), it can be interpreted as a positive valence dimension. The fourth factor accounted for 8.87% of the variance. Based on the three items that had their highest loading on this factor (curious, surpise, confused), it can be interpreted as representing an epistemic dimension.

The third solution (emotions after) pointed toward a three-factor structure that explained 58.42% of the total variance. After varimax rotation, the first factor accounted for 29.44% of the variance. Based on the eight items that had their highest loadings on this factor (sadness, disappointment, hopelessness, shame, anger, frustration, fear, anxiety), it can be interpreted as representing a negative valence dimension. The second factor accounted for 17.43% of the variance. Based on the seven items that had their highest loading on this factor (pride, happiness, relief, relaxation, gratitude, enjoyment, hopefulness) it can be interpreted as a positive valence dimension. Finally, the third factor accounted for 13.95% of the variance. Based on the four items that had their highest loading on this factor (curiosity, surprise, confusion, compassion), it appeared to be somewhat complex, but generally indicative of the epistemic dimension (with the exception of compassion). Collectively, a three-factor solution seemed to be the most consistent across time, demonstrating negative, positive, and epistemic dimensions. Rotations converged between 5 – 17 iterations. The mean Kaiser-Meyer-Olkin measure of sampling adequacy ranged from .72 – 74, indicating that the data were suitable to principal component analysis.

Discussion

The results from correlation analyses provide evidence supporting the criterion-related validity of the MEES. Specifically, emotions measured using the scale were significantly related to cognitive appraisals (concurrent validity), performance (predictive validity), and post-task self-efficacy (predictive validity). Specifically, negative emotions were negatively related to control and value appraisals, as well as negatively related to performance and post-task self-efficacy. On the other hand, positive emotions were positively related to control and value appraisals and post-task self-efficacy. These patterns are consistent with predictions according to control-value theory and provide support for the validity of the MEES, and more broadly, for the utility and generalizability of control-value theory, by demonstrating that appraisals, emotions, and performance are significantly related to one another across a variety of medical tasks that collectively represent a diverse range of skills (technical, procedural, diagnostic reasoning).

Taking this a step further, the hierarchical regression analyses expanded on these results and more directly assessed the validity of control-value theory by testing the significance of a statistical model based on this theory. The results revealed that models including control and emotions significantly predicted both performance and self-efficacy. These results demonstrated that models were significantly improved when adding the control appraisal. Although not unexpected that control would account for unique variance beyond emotions—as it is a multi-faceted construct likely influenced by prior knowledge, which has been shown to predict achievement (Hailikari et al., 2008)—what is particularly important is that after adding control, emotions remain statistically significant predictors. This suggests that the *joint* predictive value of emotions and control improves the model, rather than control alone, which provides support for the role of emotions. Future analyses could expand on these results by more directly testing the mediating role of emotions using path analyses.

Overall, the findings were largely consistent with control-value theory, although some exceptions were found. Specifically, the results revealed that of the two appraisals, control more consistently predicted emotions and performance. It may be the case that there was less variance in the value appraisal given that most participants may have considered the tasks to be highly important for professional development. Nevertheless, value was positively related to positive emotions (e.g., enjoyment, curiosity), which were related to post-task self-efficacy. This suggests that value may be more strongly tied to motivational facets of learning rather than strictly performance measures. In addition, when a task is highly valued yet the outcome is uncertain, this is hypothesized in control-value theory (Pekrun, 2006, Pekrun & Perry, 2014) to elicit anxiety, which may hamper performance. As such, the findings may represent the potentially complex relations between value and performance (i.e., it may serve as an indirect negative and positive predictor of performance depending on the type of emotions elicited).

In terms of emotions, *activating* emotions (e.g., enjoyment, anxiety) were most consistently related to appraisals and task outcomes compared to *deactivating* emotions (e.g., relaxation, boredom). This physiological activation may be a key agent in task processes that more directly impact performance (e.g., concentration, motivation). In addition, emotions experienced *during* the task were more consistently related to appraisals and task outcomes than those experienced before or after. From a measurement perspective, variations in responses over time are encouraging as they suggest that participants adhered to scale instructions and reported emotions for the targeted time point. From a theoretical standpoint, it may be the case that emotions *during* the task are more closely linked to performance as they represent the product of regulation efforts (or lack thereof) as the task is conducted.

According to control-value theory, learner characteristics and factors embedded within the broader learning environment (i.e., distal antecedents) impact emotions through their influence on cognitive appraisals (i.e., proximal antecedents) for the task at hand (Pekrun & Perry, 2014). The results from this study suggest, consistent with Study 1, that these antecedents are multifaceted.[53] The majority of participants reported that factors such as case complexity, prior knowledge, problem-solving strategies, information acquired during the task, anticipated performance, and feedback contributed to their emotions. This is also consistent with our previous work in the surgical environment (Duffy et al., 2015b), which has demonstrated through

[53] Further work is needed to determine which factors should be classified as distal antecedents and which factors should be considered components of the cognitive appraisals themselves.

retrospective interviews that participants link these factors, as well as time pressure and social dynamics (in team settings), to states of both positive and negative emotions.

With respect to exploratory analysis, results from cluster analysis suggested that participants formed three distint profiles based on whether they experienced: more intense positive emotions, more intense negative emotions, or low emotional intensity overall. This last grouping indicates that, although emotions can be used to differentiate profiles of medical trainees according to positive/negative valence, emotions may not have been activated during learning activities to a level of consciousness for a subset of individuals. It may also be the case that this low level activation represents successful emotion regulation. In terms of the structure of emotions, results from principal component analysis indicate that the dimensions linking emotions largely align with positive/negative valence, similar to cluster analyses results (with the exception of boredom, which did not appear to group with other negative emotions). In addition, the unique grouping of epistemic emotions suggests that there is another factor accounting for these emotions, which we discuss below. Given that we focused on exploratory factor analyses, further work could expand on these findings by conducting confirmatory analysis with a larger sample to directly test whether the factors align with those proposed by theory. In addition, results from these exploratory analyses should be interpreted cautiously as solutions with different numbers of clusters and dimensions may be produced in a new sample.[54]

Conclusions

The primary goal of this research was to examine the validity of an emotion scale by using control-value theory as an overarching framework. We also sought to test the relevance and generalizability of control-value theory within medical education and to explore unexamined patterns that may aid further development of self-report scales and emotion theory for medical education. This study is the first of its kind to develop and test a self-report measure of emotions specifically designed for medical education. This research also employed diverse methodologies to better understand the nature of emotions and expanded on previous work by examining states rather than traits within authentic medical learning environments rather than classrooms. Next, we discuss findings, limitations and future directions for measurement, theory, and education.

[54] Minimum samples sizes recommended for exploratory factor analysis/principal component analysis range from Subject to Variable Ratios (SVR) of 3-20 (Mundfrom et al., 2005). This study meets a SVR of 3 (20 emotion items, sample size of $N = 60$). Furthermore, there is evidence that results from smaller sample sizes are more reliable when there are higher ratios of number of variables to factors, as was the case in our results (Mundfrom et al., 2005).

Validation

In this study we have established content validity (theoretical grounding, adequate coverage of construct), face validity (scale perceived as meaningful and interpretable), criterion-related validity (correlated with theoretically related constructs), and convergent validity (aligned with interview responses). The findings provide initial evidence that the MEES is a valid self-report scale for measuring emotions across multiple authentic medical training environments and that predictions based on control-value theory are supported within these environments.

Cognitive interview responses in Study 1 indicated that the emotions presented on the MEES captured the range of emotions experienced by participants in the medical environment, although there were variations in interpretations and response choices. Retrospective interview responses in Study 1 revealed that there was strong alignment between the emotions reported though concurrent questionnaires and retrospective interviews. These results suggest that the MEES can be used to accurately capture the range and intensity of emotions experienced by medical trainees in authentic medical contexts. The results from Study 2 indicate the emotions measured using the MEES align with the conceptual understanding of the construct of emotions as described by control-value theory. In addition, the more exploratory analyses conducted in both Study 1 and 2 identified important antecedents and factor structures of emotions that can be used to inform the development or revision to scales and to enrich the control-value model.

Measurement Implications

Although the findings from this research provide evidence supporting the validity of this scale, the results also lay the groundwork for further development of emotion measures. Given that certain emotion words may be more difficult to interpret (e.g., disgust), future work could examine whether including definitions or illustrative examples in scales heightens clarity and relevance. Similarly, attaching descriptive anchors to scale intervals (particularly for the mid-way point) may lead to more consistent interpretations and coherent response patterns. Although we included emotions from multiple taxonomies (i.e., achievement, social, epistemic, basic) in the MEES, incorporating a more exhaustive range from each category would allow researchers to empirically test which emotions capture the most variance and predictive power. For instance, additional social emotions, such as envy or jealously, may be activated during group learning. Creating a long-form of MEES would allow for this kind of expansion, as well as more in-depth testing of scale reliabilities through confirmatory analyses.

It would also be worth further examining the cognitive process underlying participants' decision-making when selecting a number to represent the intensity of an emotion during the task when the scale is administered at intervals (concurrently) or after the task (retrospectively). The range of time that participants can accurately report emotions is unclear. The process of translating emotions experienced over a longer duration to self-report data may be susceptible to memory distortions. On the other hand, interrupting the task to assess emotions in the moment may also introduce biases in self-report. It is not clear how many time points are needed to accurately capture the summation of emotions experienced or whether interrupting the task at hand interferes with learning processes or gives rise to emotions that are an artifact of the methodology rather than the task (e.g., frustrated/bored due to questionnaire).

In this research we focused on emotion intensity; however, there is some evidence that suggests that the persistence or decay rate of an emotional state (rather than intensity) is an important factor in differentiating emotional experiences (D'Mello & Graesser, 2011). It may be the case that intensity scales prompt a holistic evaluation of emotions that encompasses persistence; this requires further empirical scrutiny. In addition, further exploration of non-linear patterns (e.g., moderate levels of intensity) would enrich our understanding of the role of emotions. Finally, in this research we relied solely on self-reported emotions. Although previous research has found that self-reported emotions align with behavioral measures, such as facial expression (e.g., Harley et al., 2015), future work should further test consistency between self-reported and behavioral or physiological measures of emotion. As Shapiro (2011) has stated, within medical professions in particular, cultural forces may influence whether professionals and trainees are as skilled at accurately recognizing their emotions as other populations.

Theoretical Implications

Exploratory analyses in this study revealed that the underlying component structures and clusters of emotions largely align with commonly-used (and theoretically-supported) distinctions between positive and negative emotions. However, the results also demonstrated that emotions that are more epistemic in nature group together (i.e., confusion, curiosity, surprise). This suggests that there is a unique factor that these emotions map onto apart from valence, which we expect to be predicted by control and value. What may be uniting this group of emotions is an appraisal not currently included in control-value theory: *novelty*. In our previous research (Duffy et al., 2015b) participants linked task novelty (i.e., new or unexpected events/information) to these types of emotions (curiosity, surprise) in retrospective interviews. Other researchers have also argued that these emotions are related to novelty, especially outside of traditional learning environments (e.g., art and aesthetic stimuli; see Silvia, 2005, 2009, 2010). Furthermore, this appraisal is included in Scherer's (2009) componential appraisal model of emotions (designed to capture a broader range of states beyond achievement emotions) and has emerged in principal component analysis conducted with general populations (see Fontaine & Scherer, 2013). Further research is needed examine the role of other potentially relevant appraisals.

Although most emotions we measured aligned with other states sharing the same valence and activation dimensions (e.g., positive activating emotions), certain deactivating emotions did not closely relate to one another. For instance, boredom, which is classified as a negative deactivating emotion, was not strongly related to other emotions within this category (e.g., hopelessness, disappointment, sadness) in cluster or principal component analysis. Similarly, in the correlational analyses, relaxation, which is classified as a positive deactivating emotion, was unrelated to relief, which also falls within this category. In both cases, the reason for this divergence is likely attributable to different causes of the emotion experienced by different subgroups. For instance, individuals who find the task less challenging (e.g., experts) may experience boredom or relaxation, whereas individuals who find the task highly challenging (e.g., novices) may be more likely to experience hopelessness or relief (after completion). These findings highlight exceptions to the classification of emotions based on valence and activation alone that may be accounted for by motivational and cognitive components of emotions. In a similar vein, the orthogonality of discrete emotions warrants further investigation to distinguish potentially overlapping states. For instance, it is unclear to what extent certain emotions are conceptually and experientially related (e.g., shame, disappointment). Some have suggested, for instance, that seemingly similar states, such as shame and guilt, can have different effects on medical trainees (e.g., Bynum & Goodie, 2014; Fraser & McLaughlin, 2014). There is also evidence suggesting that empathy and compassion can have different effects on burnout (Klimecki et al., 2014). Addressing these areas will help us to better understand which lexicons and taxonomies of emotions provide the most reliable accounts of emotional experiences.

In this research, we were primarily interested in testing the validity of the emotion measure and the generalizability of control-value theory across a range of authentic medical learning environments; however, there is a need to tease apart how features of the environment (e.g., performance versus learning culture, social versus individual learning, simulation versus real case) and different types of skills (e.g., technical, cognitive, procedural) relate to the types of emotions experienced and their role in learning and performance. To address this gap, in depth study of the similarities and differences between environments is needed with respect to greater standardization and larger sample sizes across environments. Future research could also examine the multi-faceted nature of control and value appraisals, by examining how valuing a task *activity* may differ from valuing a task *outcome*. Future work could also explore the temporal nature of control and value appraisals by assessing change over time in response to task events.

Educational Implications

In terms of educational implications, the results suggest that increasing trainees' perceptions of control over their ability to succeed may help to improve performance by reducing the intensity of negative emotions, at the same time as promoting positive emotions. One approach is to increase task clarity (e.g., expectations) and additional training opportunities prior to the task to better equip learners to manage challenges. During the task, scaffolds may help to decrease negative emotions. For example, within team settings such as the operating room or the deteriorating patient simulation, team members can provide support when faced with unexpected complications or cognitive overload. Within computer-based learning environments, such as BioWorld, the integration of pedagogical agents may help to enhance control by providing prompts and feedback for effective regulation. If participants feel they can improve their skills and performance during the task by managing challenges, then they may be better equipped to effectively regulate or prevent the onset of negative emotional states. Finally, the degree of complexity integrated into the design of simulations should be carefully considered in light of learners' problem-solving skills and prior knowledge. By designing training environments that support autonomy and decrease the intensity of negative emotions, by assessing trainees control appraisals during training, and by modelling effective emotion regulation (Gross, 2015), trainees may be better equipped to manage challenges and enhance performance during medical practice.

Concluding Remarks

As research moves toward state rather than trait approaches to emotion and learning, it is important to use measures that are capable of capturing a range of emotional states at the same time as remaining practical and non-intrusive for the task at hand. The use of one-word adjectives to describe an emotional state has been commonly employed in educational research examining emotions at the task level (e.g., *boredom*, Goetz et al., 2014). However, these types of measures are often questioned for their validity as it is assumed that more traditional long-form questionnaires with statement items (e.g., I feel happy that I was able to succeed on this task) are more robust. However, these types of questionnaires are typically designed to measure trait emotions and are not feasible to administer when measuring emotions in real-time at multiple instances of an activity. The results from this research, as well as other work examining the cognitive validity of emotion words (e.g., Loderer et al., under review), provide mounting evidence that these scales can provide valid measures of subjective emotional experiences.

Furthermore, by establishing the generalizability of control-value theory across three authentic medical learning environments, this research empirically tests claims that this framework can be used to inform our understanding of emotions in medical education (Artino et al., 2012). A growing body of research has found evidence supporting its utility within sport psychology (e.g., Kavussanu, Dewar, & Boardley, 2014) and primary and secondary classrooms (e.g., Pekrun & Linnenbrink-Garcia, 2014); however, the findings from our work add further support for its relevance within medical environments. As Kane (1992) and others (Karabenick et al., 2007; Mcintire & Miller 2007) have suggested, validation is an ongoing process that is strengthened by multiple studies that gather a body or chain of evidence. This research offers theoretical and methodological contributions toward this initiative.

References

Arora, S., Sevdalis, N., Aggarwal, R., Sirimanna, P., Darzi, A., & Kneebone, R. (2010). Stress impairs psychomotor performance in novice laparoscopic surgeons. *Surgical Endoscopy and Other Interventional Techniques, 24,* 2588–2593.

Artino, A. J., & Durning, S. J. (2011). It's time to explore the role of emotion in medical students' Learning. *Academic Medicine, 86*, 275–276.

Artino, A. R., Holmboe, E. S., & Durning, S. J. (2012). Control-value theory: Using achievement emotions to improve understanding of motivation, learning, and performance in medical education. *Medical Teacher, 34*, 148–160.

Artino, A. R., La Rochelle, J. S., & Durning, S. J. (2010). Second year medical students' motivational beliefs, emotions, and achievement. *Medical Education*, *44*, 1203–1212.

Artino, A. R., & Naismith, L. M. (2015). "But how do you really feel?" Measuring emotions in medical education research. *Medical Education*, *49*(2), 140–142.

Bandura, A. (2006). Guide for constructing self-efficacy scales. In F. Pajares & T. Urdan (Eds.), *Self-efficacy beliefs of adolescents* (Vol. 5, pp. 307–337). Greenwich, CT: Information Age Publishing.

Bynum, W. E., & Goodie, J. L. (2014). Shame, guilt, and the medical learner: Ignored connections and why we should care. *Medical Education, 48,* 1045–1054.

Camara, W. J., & Lane, S. (2006). A historical perspective and current views on the standards for educational and psychological testing. *Educational Measurement: Issues and Practice, 25*, 35–41.

Cannon, W. B. (1927). The James-Lange theory of emotions: A critical examination and an alternative theory. *The American Journal of Psychology 39*, 106–124.

Chew, B. H., Zain, A. M., & Hassan, F. (2013). Emotional intelligence and academic performance in first and final year medical students: A cross-sectional study. *BMC Medical Education*, *13*, 1–10.

Cook, D. A., Brydges, R., Ginsburg, S., & Hatala, R. (2015). A contemporary approach to validity arguments: A practical guide to Kane's framework. *Medical Education, 49,* 560–575.

Cook, D. A., Zendejas, B., Hamstra, S. J., Hatala, R., & Brydges, R. (2014). What counts as validity evidence? Examples and prevalence in a systematic review of simulation-based assessment. *Advances in Health Sciences Education, 19,* 233–250.

Croskerry, P., Abbass, A. A., & Wuu, A. W. (2008). How doctors feel: Affective issues in patients' safety. *The Lancet, 372,* 1205–1206.

Dehning, S., Gasperi, S., Krause, D., Meyer, S., Reiß, E., Burger, M., … Siebeck, M. (2013). Emotional and cognitive empathy in first-year medical students. *ISRN Psychiatry*, *2013*, 801530.

D'Mello, S. (2013). A selective meta-analysis on the relative incidence of discrete affective states during learning with technology. *Journal of Educational Psychology, 105,* 1082–1099.

D'Mello, S., & Graesser, A. (2011). The half-life of cognitive-affective states during complex learning. *Cognition and Emotion*, *25*, 1299–1308.

Doty, D. B., & Doty, J. R. (2012). Coronary artery bypass graft. In D. B Doty & J. R. Doty (Eds.), *Cardiac surgery: Operative technique (2nd Ed.)* (pp. 393–431). Philadelphia, PA: Elsivier.

Duffy, M. C., Lajoie, S., & Lachapelle, K. (2016). Measuring emotions in medical education: Methodological and technological advances within authentic medical learning environments. In S. Bridges, L. K. Chan, & C. E. Hmelo-Silver (Eds.), *Educational technologies in contextual learning: Research in health professions education* (pp. 181–213). Switzerland: Springer.

Dyrbye, L. N., West, C. P., Satele, D., Boone, S., Tan, L., Sloan, J., & Shanfelt, T. D. (2014). Burnout among U.S. medical students, residents, and early career physicians relative to the general U.S. population. *Academic Medicine, 89,* 443–451.

Ekman, P. (1992). An argument for basic emotions. *Cognition and Emotion*, *6*, 169.

Ellsworth, P. C. (2013). Appraisal theory: Old and new questions. *Emotion Review*, *5*, 125–131.

Ericsson, K., & Simon, H. (1993). *Protocol analysis: Verbal reports as data* (2nd ed.). Boston: MIT Press.

Fisher, A. H., & van Kleef, G. A. (2010). Where have all the people gone? A plea for including social interaction in emotion research. *Emotion Review, 2,* 208–211.

Fontaine, J. R., Scherer, K. R., & Soriano, C. (Eds.). (2013). *Components of emotional meaning: A sourcebook*. OUP Oxford.

Fontaine, J. J. R., & Scherer, K. R. (2013). The global meaning structure of the emotion domain: Investigating the complementarity of multiple perspectives on meaning. In J. J. R. Fontaine, K. R. Scherer, & C. Soriano (Eds.), *Components of emotional meaning. A sourcebook* (pp. 106–125). Oxford: Oxford University Press.

Fraser, K., & McLaughlin, K. (2014). Can we avoid the guilt of shame in medical education? *Medical Education, 48,* 1034–1042.

Gardner, D. G., Cummings, L. L., Dunham, R. B., & Pierce, J. L. (1998). Single-item versus multiple-item measurement scales: An empirical comparison. *Educational and Psychological Measurement*, *58*, 898–915.

Gauthier, G., & Lajoie, S. P. (2014). Do expert clinical teachers have a shared understanding of what constitutes a competent reasoning performance in case-based teaching? *Instructional Science, 42,* 579–594.

Goetz, T., Bieg, M., Lüdtke, O., Pekrun, R., & Hall, N. C. (2013). Do girls really experience more anxiety in mathematics? *Psychological Science*, *24*(10), 2079–2087.

Goetz, T., Frenzel, A., Hall, N. C., Nett, U. E., Pekrun, R., & Lipnevich, A. A. (2014). Types of boredom: An experience sampling approach. *Motivation and Emotion, 38,* 401–419.

Goetz, T., Frenzel, A. C., Pekrun, R., Hall, N. C., & Lüdtke, O. (2007). Between- and within-domain relations of students' academic emotions. *Journal of Educational Psychology, 99*, 715–733.

Goetz, T., Pekrun, R., & Hall, N. (2006). Academic emotions from a social-cognitive perspective: Antecedents and domain specificity of students' affect in the context of latin instruction. *The British Psychological Society, 76*, 289-308.

Gogol, K., Brunner, M., Goetz, T., Martin, R., Ugen, S., Keller, U., … Preckel, F. (2014). "My questionnaire is too long!" The assessments of motivational-affective constructs with three-item and single-item measures. *Contemporary Educational Psychology*, *39*(3), 188–205.

Greene, J. A., Yu, S. B., & Copeland, D. Z. (2014). Measuring critical components of digital literacy and their relationships with learning. *Computers & Education, 76,* 55–69.

Greene, J. A., Dellinger, K., Binbasaran Tuysuzoglu, B., & Costa, L. (2013). A two-tiered approach to analyzing self-regulated learning process data to inform the design of hypermedia learning environments. In R. Azevedo, & V. Aleven (Eds.), *International handbook of metacognition and learning technologies* (pp. 117–128). New York: Springer.

Gross, J. J. (2015). The extended process model of emotion regulation: Elaborations, applications, and future directions. *Psychological Inquiry, 26,* 130–137.

Hailikari, T., Katajavuori, N., & Lindblom-Ylanne, S. (2008). The relevance of prior knowledge in learning and instructional design. *American Journal of Pharmaceutical Education, 72,* 113–121.

Izard, C. E. (2007). Basic emotions, natural kinds, emotion schemas, and a new paradigm. *Perspectives on Psychological Science*, *2*, 260–280.

James, W., & Lange, C. G. (1922). *The emotions*. Baltimore: Williams & Wilkins Co.

Kaplan, S., Bradley, J. C., Luchman, J. N., & Haynes, D. (2009). On the role of positive and negative affectivity in job performance: A meta-analytic investigation. *Journal of Applied Psychology*, *94*, 162–176.

Karabenick, S. A., Woolley, M. E., Friedel, J. M., Ammon, B. V., Blazevski, J., Bonney, C. R., . . . Musu, L. (2007). Cognitive processing of self-report items in educational research: Do they think what we mean? *Educational Psychologist*, *42*(3), 139–151.

Kavussanu, M., Dewar, A. J., & Boardley, I. D. (2014). Achievement goals and emotions in athletes: The mediating role of challenge and threat appraisals. *Motivation and Emotion*, *38*(4), 589–599.

Klimecki, O. M., Leiberg, S., Ricard, M., & Singer, T. (2014). Differential pattern of functional brain plasticity after compassion and empathy training. *Social Cognitive and Affective Neuroscience, 9,* 873–879.

Lajoie, S. P. (2009). Developing professional expertise with a cognitive apprenticeship model: Examples from Avionics and Medicine. In K. A. Ericsson (Ed.). *Development of professional expertise: Toward measurement of expert performance and design of optimal learning environments* (pp. 61–83). New York: NY. Cambridge University Press.

Lajoie, S., Poitras, E., Naismith, L., Gauthier, G., Summerside, C., & Kazemitabar, M. (2013). Modelling domain-specific self-regulatory activities in clinical reasoning. In H. C. Lane, K. Yacef, & J. Mostow (Eds.), *International artificial intelligence and education* *proceedings* (pp. 632–635). Amsterdam: IOS Press.

LeBlanc, V. R., & Bandiera, G. W. (2007). The effects of examination stress on the performance of emergency medicine residents. *Medical Education, 41,* 556–564.

LeBlanc, V. R., MacDonald, R. D., McArthur, B., King, K., & Lepine, T. (2005). Paramedic performance in calculating drug dosages following stressful scenarios in a human patient simulator. *Prehospital Emergency Care, 9,* 439–444.

LeBlanc, V. R., McConnell, M. M., & Monteiro, S. D. (2015). Predictable chaos: a review of the effects of emotions on attention, memory and decision-making. *Advances in Health Sciences Education, 20*, 265–282.

Levine, E. L., Xu, X., Yang, L., Ispas, D., Pitariu, H. D., Bian, R., … Musat, S. (2011). Cross-national explorations of the impact of affect at work using the state-trait emotion measure: A coordinated series of studies in the United States, China, and Romania. *Human Performance, 24,* 405–442.

Loderer, K., Gentsch, K., Duffy, M. C., Pekrun, R., & Scherer, K., Meier, E., Fontaine J. R., & Soriano, C. (under review). Profiling achievement emotions: A control-value approach using the achievement emotions CoreGRID. *Motivation and Emotion*.

Marteau, T. M., & Bekker, H. (1992). The development of a six-item short-form of the state scale of the Spielberger State-Trait Anxiety Inventory (STAI). *British Journal of Clinical Psychology, 31,* 301–306.

Martin, J. A., Regehr, G., Reznick, R., Macrae, H., Murnaghan, J., Hutchison, C. & Brown, M. (1997). Objective structured assessment of technical skill (OSATS) for surgical residents. *British Journal of Surgery, 84*, 273–278.

Mauss, I. B., & Robinson, M. D. (2009). Measures of emotions: A review. *Cognition and Emotion, 23*, 209–237.

McConnell, M. M. & Eva, K. W. (2012). The role of emotion in the learning and transfer of clinical skills and knowledge. *Academic Medicine, 87,* 1316–1322.

McIntire, S. A. & Miller, L. A. (2007). *Foundations of psychological testing* (2nd ed.). Thousand Oaks, CA: Sage.

Meyers, L. S., Gamst, G., & Guarino, A. J. (2013). *Applied multivariate research: Design and interpretation* (2nd ed.). Thousand Oaks, CA: Sage.

Moors, A. (2014). Flavors of appraisal theories of emotion. *Emotion Review, 6*, 303–307.

Moors, A., Ellsworth, P. C., Scherer, K. R., & Frijda, N. H. (2013). Appraisal theories of emotion: State of the art and future development. *Emotion Review, 5*, 119–124.

Muis, K., Duffy, M., Trevors, G, & Ranellucci, J., & Foy, M. (2014). What were they thinking? Using cognitive interviewing to examine the validity of self-reported epistemic beliefs. *International Journal of Education, 2,* 17–32.

Mundfrom, D. J., Shaw, D. G., & Lu Ke, T. (2005). Minimum sample size recommendations for conducting factor analyses. *International Journal of Testing, 5,* 159–168.

Nett, U. E., Goetz, T., & Hall, N. C. (2011). Coping with boredom in school: An experience sampling perspective. *Contemporary Educational Psychology*, *36*(1), 49–59.

Pajares, F. (1996). Self-efficacy beliefs in achievement settings. *Review of Educational Research, 66,* 543–578.

Pekrun, R. (2006). The control-value theory of achievement emotions: Assumptions, corollaries, and implications for educational research and practice. *Educational Psychology Review, 18*, 315–341.

Pekrun, R., & Bühner, M. (2014). Self-report measures of academic emotion. In R. Pekrun & L. Linnenbrink-Garcia (Eds.), *International handbook of emotions in education* (pp. 561–579). New York: Francis & Taylor/Routledge.

Pekrun, R., Frenzel, A. C., Goetz, T., & Perry, R. P. (2007). The control-value theory of achievement emotions: An integrative approach to emotions in education. In P. A. Schutz & R. Pekrun (Eds.), *Emotion in education* (pp. 13–36). San Diego, CA: Academic Press.

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,* 36–48.

Pekrun, R., Goetz, T., & Perry, R. P. (2005). Achievement Emotions Questionnaire (AEQ): User's manual. *Munich, Germany: University of Munich, Department of Psychology*.

Pekrun, R., & Linnenbrink-Garcia, L. (Eds.). (2014). *International handbook of emotions in education*. New York: Taylor & Francis/Routledge.

Pekrun, R., & Perry P. P. (2014). Control-value theory of achievement emotions. In R. Pekrun & L. Linnenbrink-Garcia (Eds.), *International handbook of emotions in education* (pp. 120–141). New York: Routledge.

Pekrun, R., & Stephens, E. J. (2012). Academic emotions. In K. R. Harris, S. Graham, T. Urdan, S. Graham, J. M. Royer, & M. Zeidner (Eds.), *APA educational psychology handbook* (Vol. 2, pp. 3–31). Washington, DC: American Psychological Association

Pintrich, P. R. (2000). The role of goal orientation in self-regulated learning. In M. Boekaerts, P. Pintrich & M. Zeidner (Eds.), *Handbook of self-regulation* (pp. 451–502). San Diego, CA: Academic Press.

Pintrich, P. (2004). A conceptual framework for assessing motivation and self-regulated learning in college students. *Educational Psychology Review, 16,* 385–407.

Posner, J., Russell, J. A., & Peterson, B. S. (2005). The circumplex model of affect: An integrative approach to affective neuroscience, cognitive development, and psychopathology. *Developmental Psychopathology, 17,* 715–734.

Reznick, R., Regehr, G., MacRae, H., Martin, J., & McCulloch, W. (1997). Testing technical skill via an innovative "bench station" examination. *American Journal of Surgery. 173*(3), 226–230.

Scherer, K. (2005). What are emotions? And how can they be measured? *Social Science Information, 44*, 695–729.

Scherer, K. R., Shuman, V., Fontaine, J. R. J., & Soriano, C. (2013). The GRID meets the wheel: Assessing emotional feeling via self-report. In J. R. J. Fontaine, K. R. Scherer, & C. Soriano (Eds.), *Components of emotional meaning: A sourcebook* (pp. 281–298). Oxford, United Kingdom: Oxford University Press.

Shuman, V., & Scherer, K. R. (2014). Concepts and structure of emotions. In R. Pekrun & L. Linnenbrink-Garcia (Eds.), *International handbook of emotions in education* (pp. 13–35). New York: Routledge.

Scherer, K. R. (2009). The dynamic architecture of emotion: Evidence for the component process model. *Cognition and Emotion, 23,* 1307–1351.

Silvia, P. J. (2005). What is interesting? Exploring the appraisal structure of interest. *Emotion*, *5*, 89–102.

Silvia, P. J. (2009). Looking past pleasure: Anger, confusion, disgust, pride, surprise, and other unusual aesthetic emotions. *Psychology of Aesthetic, Creativity, and the Arts, 3,* 48–51.

Silvia, P. J. (2010). Confusion and interest: The role of knowledge emotions in aesthetic experience. *Psychology of Aesthetics, Creativity, and the Arts, 4,* 75–80.

Spielberger, C. D. (1989). *State-trait anxiety inventory: Bibliography* (2nd ed.). Palo Alto, CA: Consulting Psychologists Press.

Spielberger, C. D., Gorsuch, R. L., Lushene, R., Vagg, P. R., & Jacobs, G. A. (1983). *Manual for the state-trait anxiety inventory*. Palo Alto, CA: Consulting Psychologists Press.

Sullivan, G. M. (2011). Getting off the “gold standard”: Randomized controlled trials and education research. *Journal of Graduate Medical Education, 3,* 285–289.

Turner, J. C., & Trucano, M. (2014) Measuring situated emotion. In R. Pekrun & L. Linnenbrink-Garcia (Eds.), *International handbook of emotions in education*. New York: Routledge.

Watson, D., & Clark, L. A. (1999). *The PANAS-X: Manual for the positive and negative affect schedule-expanded form*. University of Iowa.

Watson, D., Clark, L. A., & Tellegen, A. (1988). Development and validation of brief measures of positive and negative affect: The PANAS scales. *Journal of Personality and Social Psychology*, *54*, 1063–1070.

Weiner, B. (2007). Examining emotional diversity on the classroom: An attribution theorist considers the moral emotions. In P. A. Shutz & R. Pekrun (Eds.), *Emotion in education* (pp. 73-88). San Diego, CA: Academic Press.

Wigfield, A., & Eccles, J. S. (2000). Expectancy-value theory of achievement motivation. *Contemporary Educational Psychology, 25,* 68–81.

Wiseman, J., & Snell, L. (2008). The deteriorating patient: A realistic but “low-tech” simulation of emergency decision-making. *The Clinical Teacher*, *5*, 93–97.

Wiseman, J. & Labelle, M. (2015, March). *Early recognition of and response to acute deterioration (ERRAD): A theory-based method for teaching medical undergraduates acute care thinking*. Presented at the Canadian Conference on Medical Education.

Yeung, Yee, K., & Ruzzo, W. L. (2000). *An empirical study on principal component analysis for clustering gene expression data* (Technical Report UW-CSE-2000-11-03). Retrieved from University of Washington website: http://faculty.washington.edu/kayee/pca/pca.pdf.

Table 14
Sample Cognitive and Retrospective Interview Questions

Interview type	Sample interview questions
Retrospective interview	Thinking back to immediately *before the case began/during the case/immediately after the case*, can you tell me what emotions you remember feeling? Were you feeling any positive emotions? Any negative emotions? What types of thoughts did you have while experiencing these emotions? Did you express these emotions at all? If so, how? Did you feel more tense or relaxed? What were the causes of these emotions?
Cognitive interview	Did you feel the questionnaire captured an adequate range of emotions? Were there any emotions that you experienced but did not see listed on the questionnaire? Were there any emotions that seemed irrelevant? Were there any emotions that were unclear or difficult to interpret? Did you feel that the questionnaire was administered at a frequency that accurately captured your emotions? When you rated the emotion, did you think about how strongly you felt the emotion or how frequently you experienced it? What did it mean when you selected “moderate” as the mid-way point on the scale?

Note. Different stems were used to prompt participants to recall their emotional experience in relation to specific time points (before, during after the task) in a sequential order. Each time point was paired with follow-up questions that asked participants to elaborate on the nature of their emotional experiences.

Table 15
Emotion Coding Scheme for Interviews

Code Type	Description	Example
Emotion		
[B.EM+ –]	Before emotion (positive/negative)	curious, excited, happy, relief, relaxed, confident, proud confused, frustrated, anxious, tense, nervous
[D.EM+ –]	During emotion (positive/negative)	curious, excited, happy, relief, relaxed, confident, proud confused, frustrated, anxious, tense, nervous
[A.EM+ –]	After emotion (positive/negative)	curious, excited, interested, happy, relief, relaxed, confident, proud confused, frustrated, anxious, tense, nervous
Antecedent		
[ANT.+ – T]	Antecedent lab test result (positive/negative)	"The urinary catecholamines were sky high I was like 'yes!' and I knew the diagnosis from there"
[ANT. + – FB]	Antecedent feedback/expert summary (positive/negative)	"I was happy when I first saw the feedback because I had the right diagnosis"
[ANT. + – Task]	Antecedent task/learning environment (positive/negative)	"I was curious in anticipation for what the case would be."
[ANT. + – PP]	Antecedent past performance (positive/negative)	"I didn't do as well on the last case so I was pretty nervous"
[ANT. + – FP]	Antecedent future performance (positive/negative)	"I am also curious as to how well I am going to do in this given task."
[ANT. + – CP]	Antecedent current performance (positive/negative)	"I was frustrated that I made mistakes"
[ANT. + – KN]	Antecedent possess/lack knowledge (positive/negative)	"I was disappointed with my knowledge in that particular case"

Expression		
[EXP.+ – PYS]	Expression physiological (increase/decrease)	Sweating, heart rate (increase/decrease)
[EXP. + – BEH]	Expression behavioural (increase/decrease)	Body language (e.g., fidgeting), facial expression (e.g., smiling,), vocal features (e.g., shaky voice)
[EXP. + – MTV]	Expression motivational (increase/decrease)	Interest, persistence, effort, drive
[EXP. + – COG]	Expression cognitive (increase/decrease)	Attending, reflecting, evaluating

Table 16
Top Five Most Intensely Experienced Positive and Negative Emotions across Time

Time	Positive	*M* (*SD*)	Negative	*M* (*SD*)
Before	Curious	3.9 (0.6)	Anxious	2.2 (0.9)
	Enjoyment	3.6 (0.7)	Afraid	1.9 (0.9)
	Hopeful	3.4 (0.8)	Bored	1.7 (1.0)
	Relaxed	3.2 (0.9)	Confused	1.6 (.8)
	Compassionate	2.9 (0.8)	Frustrated	1.5 (0.7)
During	Curious	3.7 (0.8)	Confused	2.4 (1.2)
	Enjoyment	3.7 (0.9)	Anxious	2.2 (1.0)
	Hopeful	3.3 (1.2)	Frustrated	1.8 (1.0)
	Compassionate	2.9 (1.3)	Disappointed	1.6 (0.8)
	Relaxed	2.8 (1.2)	Afraid	1.5 (0.6)
After	Enjoyment	3.3 (0.9)	Confused	1.9 (0.8)
	Relaxed	3.1 (1.1)	Disappointed	1.7 (0.9)
	Curious	3.1 (1.1)	Frustrated	1.6 (0.7)
	Hopeful	3.0 (1.1)	Ashamed	1.5 (0.6)
	Proud	2.9 (1.3)	Anxious	1.5 (0.6)

Table 17
Retrospective Interview Results (Group-level Alignment)

Phase	Category	Sample quotation	Overall pattern
Before	**Curiosity/interest (*n* = 8)**	"Curious… I wondered what this patient had"	57% of participants reported feeling curious before the task 29% reported feeling anxious before the task.
	Happiness/ excitement (*n* = 6)	"[Feeling] excited or looking forward to doing this case"	
	Hopeful (*n* = 4)	"I was hoping that it wouldn't be so difficult."	
	Pride/confidence	"Proud because I had done well on my first case"	
	Relaxation	"I was relaxed. Yeah I think I was kind of just ready to take my time"	
	Anxiety/stress/ nervousness (*n* = 4)	"Anxiety because it's a new case"	
	Confusion/ uncertainty (*n* = 4)	"I was feeling a bit confused because I wasn't really sure if I would be able to diagnose anything at all."	
During	**Happiness/ excitement/ enjoyment (*n* = 11)**	"I was kind of just excited to try to figure it out"	73% of participants reported feeling happy/enjoyment/excited during the task. 33% of participants reported feeling confused/uncertain/frustrated during the task
	Relief/relaxation/ (*n* = 7)	"I remembered feeling more and more relieved as I went through it because it seemed a lot easier...."	
	Pride/certain/ satisfaction (*n* = 7)	"I was pretty confident that I knew what it was going to be."	
	Curious (*n* = 4)	"I was curious to see if I was right"	
	Hopeful (*n* = 1)	"Hopeful that I was on the right track"	
	Confusion/ uncertainty/ frustration (*n* = 5)	"I remember being a bit confused… cause it was a more complex case."	
	Anxiety/fear (*n* = 3)	"It's also a fear associated with that and the fact that I'm not able to explain why this test gave me that kind of results."	

Phase	Category	Sample quotation	Overall pattern
After	**Happiness/glad (*n* = 10)**	**“**I was glad that I got the right diagnosis and most of the symptoms were right.”	66% of participants reported feeling happy/glad after the task 27% of participants reported feeling confused/frustrated after the task.
	Pride/satisfaction/ confidence (*n* = 6)	“I was proud of myself that I could match with an expert.”	
	Relief (*n* = 4)	“I was relieved that I had the correct diagnosis."	
	Curiosity/intrigue (*n* = 3)	“And then was curious about you know, if I had the right reasoning or not"	
	Grateful (*n* = 2)	“I feel grateful I mean in a way to know that I’m still in a learning process."	
	Hopeful (*n* = 2)	“Hopeful…I’m thinking forward in terms of like me in clinical settings"	
	Confusion/ uncertainty/ frustration (*n* = 4)	“[confused] Because I missed things [evidence/information]”	
	Disappointment/ guilt (*n* = 3)	“Disappointed that I missed a test”	
	Anxiety/fear (*n* = 2)	“Fear... that I had completely missed something that’s so obvious”	
	Surprise/shock (*n* = 1)	“[Shocked because] I saw the diagnosis”	

Note. Bolded cells indicate a match between interview and MEES responses at each time point. A match was identified when the most frequently reported positive and negative emotion in the interview (based on group-level percentage) aligned with the highest ranked positive and negative emotions in the MEES (based on group-level mean). We did not use intensity for this coding as participants reported in interviews the one emotion that they felt *most* intensely and often this only included in a positive state. Our goal was to examine alignment between interviews and the MEES at the group level for both positive and negative states at each phase.

Table 18
Sample Alignment between Individual Interview Responses and MEES

	Before		During		After	
Participant (N = 15)	MEES	Interview	MEES	Interview	MEES	Interview
1	Happy (4)	Excited	**Confused (5)**	**Confused** Frustrated	Curious (5)	Surprised Relieved
2	Curious (4)	Curious	Confused (4)	**Confused**	Hopeful (4)	Disappoin ted
3	Curious (4)	Curious	Confused (4)	Confused	**Happy (4)**	Happy
4	**Curious (5)**	Curious	Happy (5)	**Happy**	**Relaxed (5)**	**Relaxed**
5	**Hopeful**	**Hopeful**	Enjoymen t (4.5)	**Excited** Proud	**Proud (4)**	**Proud**
6	**Proud (4)**	**Proud**	**Curious (4)**	Curious	**Proud (5)**	**Proud**
7	**Curious (4)**	Curious	**Relieved (4)**	Relieved	**Happy (3)**	Happy

Note. Bolded cells indicate a match between interview response and MEES and the number in brackets indicates the intensity rating of the emotion from the MEES. Interview responses were selected based on the emotion the participant reported experiencing most intensely. For the MEES during, an average score was calculated to identify the overall most intensely experienced emotion. A match was identified if the most intense emotion(s) aligned with at least one of the most intense emotions identified on the MEES (in the case of two or more emotions sharing the same mean score) for the corresponding phase of the task.

Table 19
Cognitive Interview Results

Question	Category	Sample Quotation	Overall Pattern
Are there any emotions you experienced but were not in the questionnaire?	Excitement	"Exciting wasn't on there"	Most participants reported that the questionnaire was not missing any emotions. Excitement and embarrassment were two notable exceptions.
	Embarrassed (n = 1)	"I wasn't ashamed of not ordering tests I was more like a little embarrassed"	
Were there any emotions that the seemed irrelevant?	Disgust (n = 9)	"Disgusted, I don't see why someone will be disgusted relating to the task"	Participants found disgust and compassion to be the most irrelevant emotions. Shame and anger were considered too intense in some cases
	Compassion (n = 8)	"It's hard to feel compassionate when you don't have a real person in front of you"	
	Shame (n = 4)	"Yes you're disappointed, but ashamed, no"	
	Anger (n = 2)	"[Anger], it's very extreme"	
Were there any emotions that were difficult to interpret?	Neutral (n = 7)	"Neutral means I'm not being at all stimulated, so it wasn't even applicable in that case"	Participants found neutral difficult to interpret.
What does moderate on the scale represent for you?	An emotion experienced, but not intensely (n = 8)	"It meant like neither, just kind of like in the middle"	Interpretation of Likert scale mid-way point varied.
	A baseline (n = 4)	"I think I used moderate just when I was sort of baseline"	

Question	Category	Sample Quotation	Overall Pattern
	Depends on the emotion (*n* = 2)	"'Moderate' was generally when I was, sort of just my baseline for positive emotions. For negative emotions I would put a 1,'Not at all'"	
	When uncertain of emotion felt (*n* = 3)	"I put moderate for a lot of things I wasn't too sure about"	
Did you report frequency or intensity of an emotion?	Intensity (*n* = 12)	"How intensely or strongly an emotion was felt"	Most participants rated intensity of emotion over frequency
	Frequency (*n* = 3)	"How often the emotion was felt"	

Note. Responses of less than two participants (*n* = 1) were excluded from the irrelevant emotions category as we were aiming to identify emotions that may be considered irrelevant at the group level; due to a technical recording issue, one participant's cognitive interviewing was not completed; their responses could not be included in calculations

Table 20
Features of Three Authentic Learning Environments

Features	Learning environments		
	Surgical environment	BioWorld environment	Deteriorating patient environment
Instructional design	Training in clinical practice through apprenticeship model, deliberate practice	Computer-based learning environment that provides virtual cases and (meta)cognitive scaffolds	Low-tech role play simulation facilitated by tutor with instructional lectures and debriefing
Learning goals	Technical skills	Diagnostic reasoning skills	Procedural knowledge
Social dynamics	Team (shared goal solved together)	Individual (personal goal solved independently)	Group (shared goal solved independently)
Fidelity	High	Moderate	Moderate
Feedback	During/after task	During/after task	During/after task

Table 21
Descriptive Statistics for Appraisals and Emotions within Learning Environments

	Surgical environment	BioWorld environment	DP simulation environment
	M (*SD*)	*M* (*SD*)	*M* (*SD*)
Variables			
Control appraisal	4.00 (.89)[a]	3.00 (1.14)[b]	3.55 (.79)
Value appraisal	4.25 (.82)	3.96 (1.12)[a]	4.70 (.57)[b]
Positive activating emotions	3.06 (.66)	2.71 (.77)	2.89 (.57)
Positive deactivating emotions	2.84 (.60)	2.72 (.78)	2.46 (.68)
Negative activating emotions	1.73 (.72)	1.99 (.70)[a]	1.50 (.32)[b]
Negative deactivating emotions	1.42 (.59)[a]	1.88 (.69)[b]	1.17 (.22)[b]
Epistemic emotions	1.88 (.66)[a]	2.84 (.65)[b]	2.20 (.50)[b]
Achievement emotions	2.30 (.31)[a]	2.27 (.44)[a]	1.94 (.37)[b]
Basic emotions	1.82 (.40)	1.83 (.52)[a]	1.55 (.19)[b]
Social emotions	2.88 (.98)	2.42 (.87)	2.78 (.68)

Note. DP refers to deteriorating patient. 1-5 Likert scale for all variables. Superscript letters that differ in the same row indicate statistically significant differences in means for post-hoc analyses ($p < .05$). ANOVAs revealed significant differences between environments for the control appraisal, $F(2, 57) = 5.33$, $p < .01$, and value appraisal, $F(2, 57) = 3.79$, $p < .05$. In terms of emotion valence, MANOVAs revealed significantly differences between environments for positive emotions, $F(4, 112) = 2.84$, $p < .05$, and negative emotions, $F(4, 112) = 4.56$, $p < .01$. In terms of emotion taxonomies, ANOVAs revealed significant differences between environments for epistemic emotions, $F(2, 57) = 13.38$, $p < .001$, achievement emoions, $F(2, 56) = 5.22$, $p < .01$, and basic emotions, $F(2, 57) = 3.15$, $p = .051$.

Table 22
Correlations between Appraisals, Emotions, Outcomes (Surgical Environment)

	1	2	3	4	5	6	7	8
1. Control								
2. Value	-.06							
3. Positive Act.	.28	.35						
4. Positive Deact.	.35	.24	.71**					
5. Negative Act.	-.46*	.20	-.13	-.48*				
6. Negative Deact.	-.42	.10	-.33	-.59**	.80**			
7. Performance	.46*	.06	-.10	-.11	-.31	.07		
8. Self-efficacy	.26	.21	.15	.19	-.47*	-.08	.71**	

Note. One-tailed test; *p < .05; **p < .01; positive act. emotions = positive activating emotions; positive deact. emotions = positive deactivating emotions; negative act. emotions = negative activating emotions; negative deact. emotions = negative deactivating emotions.

Table 23
Correlations between Appraisals, Emotions, Outcomes (BioWorld Environment)

	1	2	3	4	5	6	7	8
1. Control								
2. Value	.58**							
3. Positive Act.	.44*	.19						
4. Positive Deact.	.47*	.28	.78**					
5. Negative Act.	-.30	.14	-.22	-.31				
6. Negative Deact.	-.33	-.06	-.28	-.27	.81**			
7. Performance	.42*	.12	.32	.19	-.41*	-.44*		
8. Self-efficacy	.22	.07	.71**	.67**	-.44*	-.46*	.29	

Note. One-tailed test; *p < .05; **p < .01; positive act. emotions = positive activating emotions; positive deact. emotions = positive deactivating emotions; negative act. emotions = negative activating emotions; negative deact. emotions = negative deactivating emotions.

Table 24
Correlations between Appraisals, Emotions, Outcomes (Deteriorating Patient Environment)

	1	2	3	4	5	6	7	8
1. Control								
2. Value	-.09							
3. Positive Act.	.28	.21						
4. Positive Deac.	.54**	-.10	.68**					
5. Negative Act.	-.25	.46*	.39*	-.12				
6. Negative Deac.	.22	.15	.50*	.40*	.30			
7. Performance	.47*	-.16	-.09	.07	-.63**	-.26		
8. Self-efficacy	.72**	-.01	.52*	.68**	-.18	.28	.39*	

Note. One-tailed test; * p < .05; ** p < .01; positive act. emotions = positive activating emotions; positive deact. emotions = positive deactivating emotions; negative act. emotions = negative activating emotions; negative deact. emotions = negative deactivating emotions.

Table 25
Hierarchical Multiple Regression Analyses

	Performance Model (negative emotions)		Self-efficacy Model (positive emotions)	
Predictor	Adjusted R^2	β	Adjusted R^2	β
Step 1	.19**		.27**	
Act. emotions before		.09		-.06
Act. emotions during		-.72**		.49*
Deact. emotions before		-.02		.06
Deact. emotions during		-.29		-.01
Step 2	.25**		.34**	
Act. emotions before		.13		-.05
Act. emotions during		-.66**		.52**
Deact. emotions before		.04		-.05
Deact. emotions during		.29		-.00
Control appraisal		.30*		.32*

Note. Act. emotions = activating emotions; Deact. emotions = deactivating emotions; Model 1 (performance) contained negative activating and deactivating emotions (step 1) and control appraisal (step 2); Model 2 contained positive activating and deactivating emotions (step 1) and control appraisal (step 2); *$p < .05$; **$p < .01$

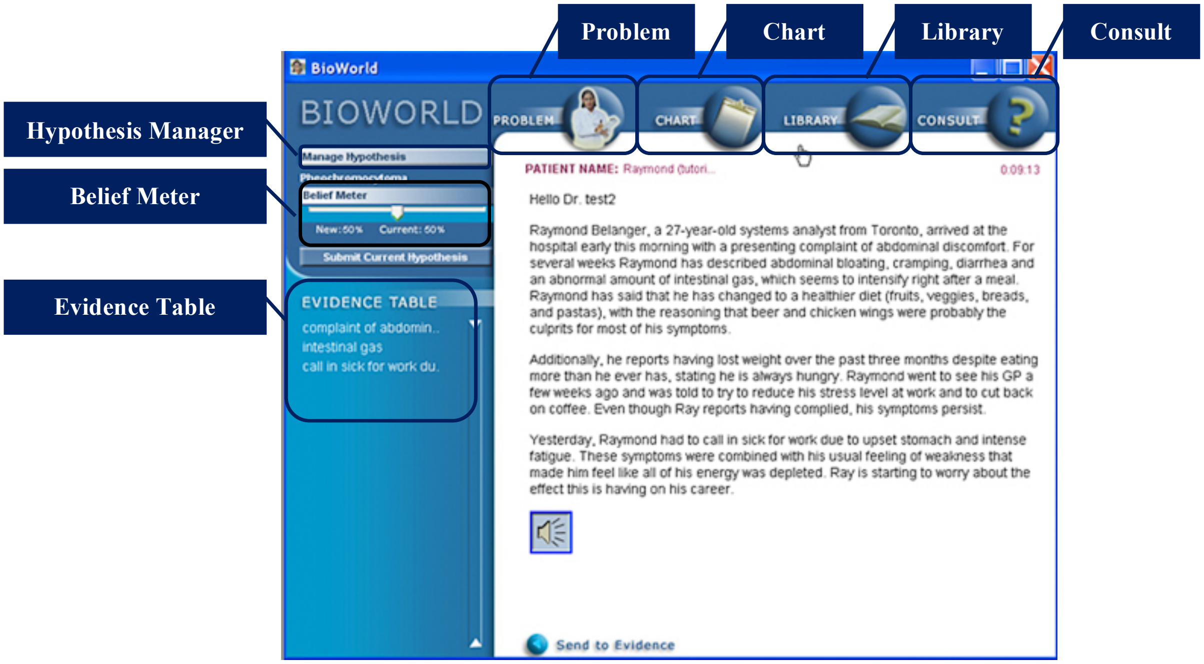


Figure 9. BioWorld environment.

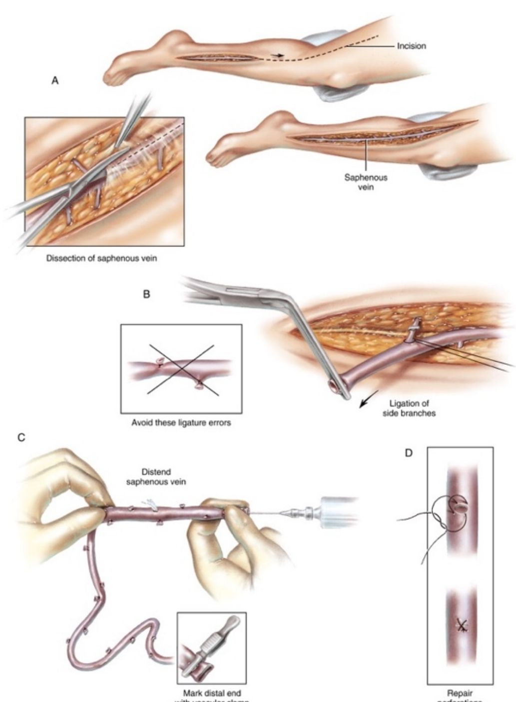


Figure 10. Open saphenous vein harvest procedure.

Figure from "Coronary Artery Bypass Graft," by D. B Doty & J. R. Doty (Eds.), *Cardiac surgery: Operative technique (2nd Ed.)* (p. 397), 2012, London, England: Elsivier. Copyright 2012 by Elsivier. Reprinted with permission.

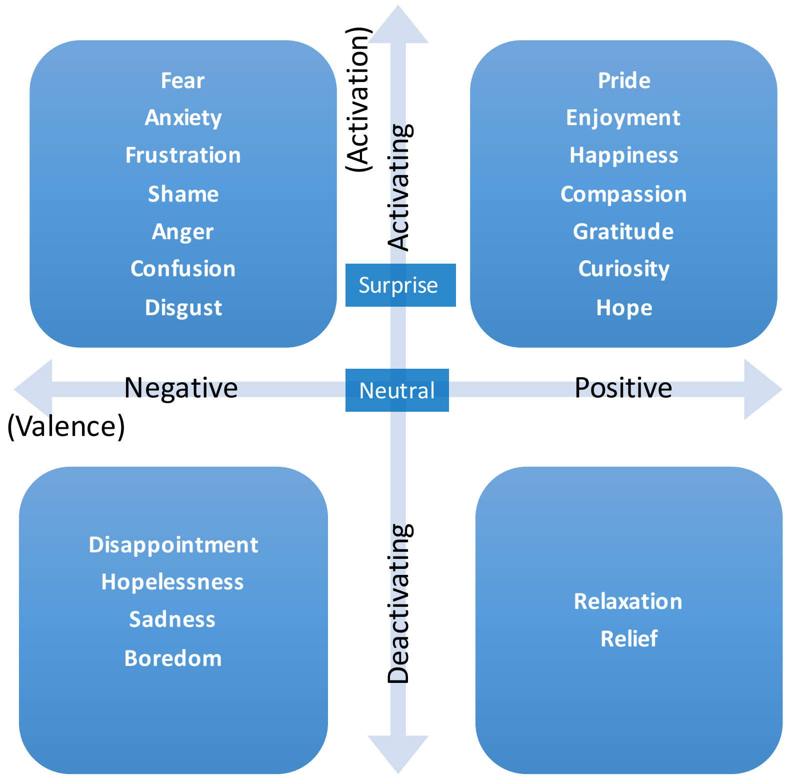


Figure 11. Emotions from the MEES classified according to valence and activation
Note. Neutral and disgust were exluded from analyses in Study 2 based on results of Study 1.

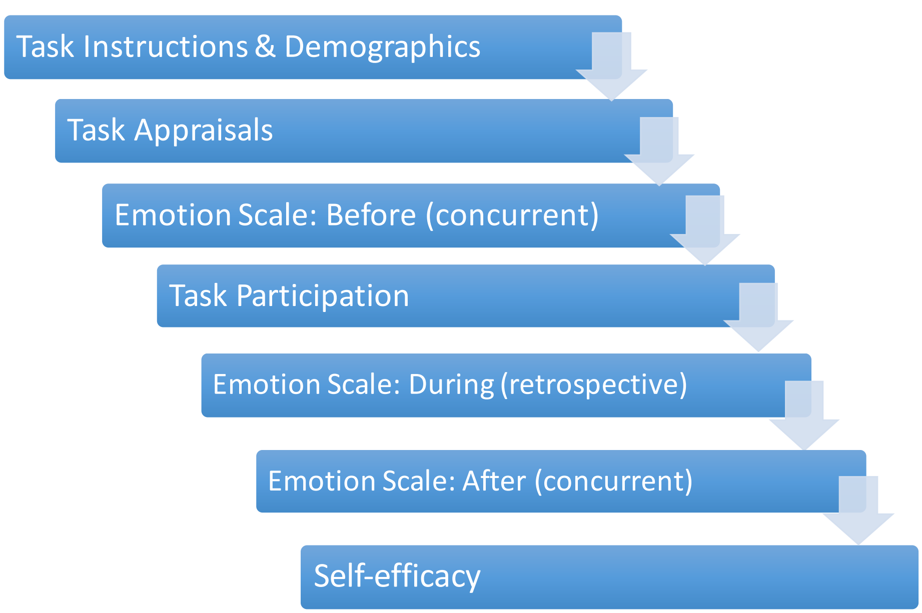


Figure 12. Research protocol for all environments.

Appendix G

Manual for Medial Education Emotion Scale

Background/Design:

- The Medical Education Emotion Scale (MEES) includes a range of single-item discrete emotion terms that can be grouped into 4 sub-scales based on taxonomy: social (e.g., compassion, gratitude), achievement (e.g., pride, shame), basic (e.g., happiness, sadness), and epistemic/knowledge emotions (curiosity, confusion). Alternatively, emotions can be grouped according to valence and activation: positive activating, positive deactivating, negative activating, negative deactivating.
- Emotions included in this scale were selected specifically for their relevance within medical education. The scale is designed to measure state emotions and the instructions can be tailored for different time points to create versions that can be administered *before*, *during*, and *after* a task or activity to examine change over time. We have opted to administer the *during* questionnaire using a time-based method (e.g., every 10 minutes for 30 minute tasks; on average twice during an activity). However, it can also be administered using an event-based method. We have also administered it for activities using a retrospective method (see below).
-

Method of Administration:

- We have administered the MEES both concurrently and retrospectively for emotions experienced *during* an activity. By modifying the instructions participants can be directed to report the emotions they are currently feeling (concurrent), or to reflect back on emotions experienced during an activity (retrospective). The wording listed in the scale is the concurrent method. However, the retrospective approach is particularly useful if there is concern that interrupting a task to administer a questionnaire would interfere with performance or place constraints on cognitive load. We typically inform participants in advance of the method we are using and that we will be administering questionnaires to examine their emotions at various time points. Sample phrasing is provided below:
  Concurrent instructions: Emotions **DURING** the activity: Using the scale below, indicate how you currently feel *keeping in mind the activity you are currently working on*. For each emotion, please indicate the strength of that emotion by selecting the number that best describes the intensity of your emotion.
  Retrospective instructions: Emotions **DURING** the activity: Keeping in mind the activity you just completed, indicate how you felt *during the activity*. For each emotion, please indicate the strength of that emotion by selecting the number that best describes the intensity of your emotion.

Emotions **BEFORE** the activity: Using the scale below, indicate how you currently feel *keeping in mind the activity you are about to begin*. For each emotion, please indicate the strength of that emotion by selecting the number that best describes the intensity of your emotion.

	1 **Not at all**	**2** **Very little**	**3** **Moderate**	**4** **Strong**	**5** **Very strong**
Confused	○	○	○	○	○
Hopeful	○	○	○	○	○
Bored	○	○	○	○	○
Proud	○	○	○	○	○
Sad	○	○	○	○	○
Anxious	○	○	○	○	○
Happy	○	○	○	○	○
Frustrated	○	○	○	○	○
Hopeless	○	○	○	○	○
Enjoyment	○	○	○	○	○
Ashamed	○	○	○	○	○
Neutral	○	○	○	○	○
Compassionate	○	○	○	○	○
Surprised	○	○	○	○	○
Curious	○	○	○	○	○
Afraid	○	○	○	○	○
Grateful	○	○	○	○	○
Disappointed	○	○	○	○	○
Relieved	○	○	○	○	○
Angry	○	○	○	○	○
Relaxed	○	○	○	○	○
Disgusted	○	○	○	○	○

--Continued on next page--

Emotions **DURING** the activity: Using the scale below, indicate how you currently feel *keeping in mind the activity you are currently working on*. For each emotion, please indicate the strength of that emotion by selecting the number that best describes the intensity of your emotion.

	1 Not at all	**2 Very little**	**3 Moderate**	**4 Strong**	**5 Very strong**
Confused	○	○	○	○	○
Hopeful	○	○	○	○	○
Bored	○	○	○	○	○
Proud	○	○	○	○	○
Sad	○	○	○	○	○
Anxious	○	○	○	○	○
Happy	○	○	○	○	○
Frustrated	○	○	○	○	○
Hopeless	○	○	○	○	○
Enjoyment	○	○	○	○	○
Ashamed	○	○	○	○	○
Neutral	○	○	○	○	○
Compassionate	○	○	○	○	○
Surprised	○	○	○	○	○
Curious	○	○	○	○	○
Afraid	○	○	○	○	○
Grateful	○	○	○	○	○
Disappointed	○	○	○	○	○
Relieved	○	○	○	○	○
Angry	○	○	○	○	○
Relaxed	○	○	○	○	○
Disgusted	○	○	○	○	○

--Continued on next page--

Emotions **AFTER** the activity: Using the scale below, indicate how you currently feel *now that the activity is completed*. For each emotion, please indicate the strength of that emotion by selecting the number that best describes the intensity of your emotion.

	1 Not at all	**2 Very little**	**3 Moderate**	**4 Strong**	**5 Very strong**
Confused	○	○	○	○	○
Hopeful	○	○	○	○	○
Bored	○	○	○	○	○
Proud	○	○	○	○	○
Sad	○	○	○	○	○
Anxious	○	○	○	○	○
Happy	○	○	○	○	○
Frustrated	○	○	○	○	○
Hopeless	○	○	○	○	○
Enjoyment	○	○	○	○	○
Ashamed	○	○	○	○	○
Neutral	○	○	○	○	○
Compassionate	○	○	○	○	○
Surprised	○	○	○	○	○
Curious	○	○	○	○	○
Afraid	○	○	○	○	○
Grateful	○	○	○	○	○
Disappointed	○	○	○	○	○
Relieved	○	○	○	○	○
Angry	○	○	○	○	○
Relaxed	○	○	○	○	○
Disgusted	○	○	○	○	○

--Continued on next page--

Appendix H
Retrospective/Cognitive Interview Guide

Instructions Part 1: *I would like to ask you a few questions about the emotions you experienced throughout the diagnostic reasoning task you just completed in BioWorld. I want to see how much you remember about what you were feeling before, during, and after you finished interacting with this specific case. I am interested in what you actually remember about what you were feeling rather than what you think you must have been feeling. I would like you to tell about your feelings in the sequence in which they occurred. Please tell me if you are uncertain about any of your memories.*

Questions Part 1

- Thinking back to *immediately before* the case began, can you tell me what emotions you remember feeling?
    - Follow-up prompt: Were you feeling any positive emotions? Any negative emotions? Which emotions did you experience most intensely? Did you express these emotions at all? If so, how? Were you feeling more tense or relaxed? What were the causes or focus of these emotions?
- Thinking back to *during* the case, while you were engaging in the diagnostic reasoning task, can you tell me what emotions you remember feeling?
    - Follow-up prompt: Were you feeling any positive emotions? Any negative emotions? Which emotions did you experience most intensely? Did you express these emotions at all? Were you feeling more tense or relaxed? What were the causes or focus of these emotions?
- Thinking back to *immediately after* the case ended, while you were engaging in the diagnostic reasoning task, can you tell me what emotions you remember feeling?
    - Follow-up prompt: Were you feeling any positive emotions? Any negative emotions? Which emotions did you experience most intensely? Did you express these emotions at all? What were the causes or focus of these emotions?

Instructions Part 2: *I would now like to ask you your thoughts about the emotions that were presented on the questionnaire and how you interpreted and responded to them.*

Questions Part 2

- Did you feel that the emotions in the questionnaire captured the range of feelings that you experienced within BioWorld?
    - **Show them copy of questionnaire for follow-up prompts (below)**
    - Follow-up prompts: Were there any emotions that you were felt but did not see listed on the questionnaire? Were there any emotions that did not seem relevant? Were there any that that seemed unclear or difficult to interpret?
- When you rated the intensity of emotion, did you think about how strongly you felt an emotion or how frequently you experienced it? What does it mean when you select “moderate” the mid-way point on the scale?
- Did you respond to items based on the emotions you were currently feeling or based on the emotions you were feeling several minutes before the questionnaire was administered? Did you feel that the questionnaire was administered at a frequency that accurately captured your emotions? Did you feel it was too much? No enough?

Appendix I
Global Rating Scale for Technical Performance

GLOBAL RATING SCALE OF OPERATIVE PERFORMANCE

Please circle the number corresponding to the candidate's performance in each category, **irrespective of training level.**

Respect for Tissue:

1	2	3	4	5
Frequently used unnecessary force on tissue or caused damage by inappropriate use of instruments		Careful handling of tissue but occasionally caused inadvertent damage		Consistently handled tissues appropriately with minimal damage

Time and Motion:

1	2	3	4	5
Many unnecessary moves		Efficient time/motion but some unnecessary moves		Clear economy of movement and maximum efficiency

Instrument Handling:

1	2	3	4	5
Repeatedly makes tentative or awkward moves with instruments by inappropriate use of instruments		Competent use of instruments but occasionally appeared stiff or awkward		Fluid moves with instruments and no awkwardness

Knowledge of Instrument:

1	2	3	4	5
Frequently asked for wrong instrument or used inappropriate instrument		Knew names of most instruments and used appropriate instrument		Obviously familiar with the instruments and their names

Flow of Operation:

1	2	3	4	5
Frequently stopped operating and seemed unsure of next move		Demonstrated some forward planning with reasonable progression of procedure		Obviously planned course of operation with effortless flow from one move to the next

Use of Assistants:

1	2	3	4	5
Consistently placed assistants poorly or failed to use assistants		Appropriate use of assistants most of the time		Strategically used assistants to the best advantage at all times

Knowledge of Specific Procedure:

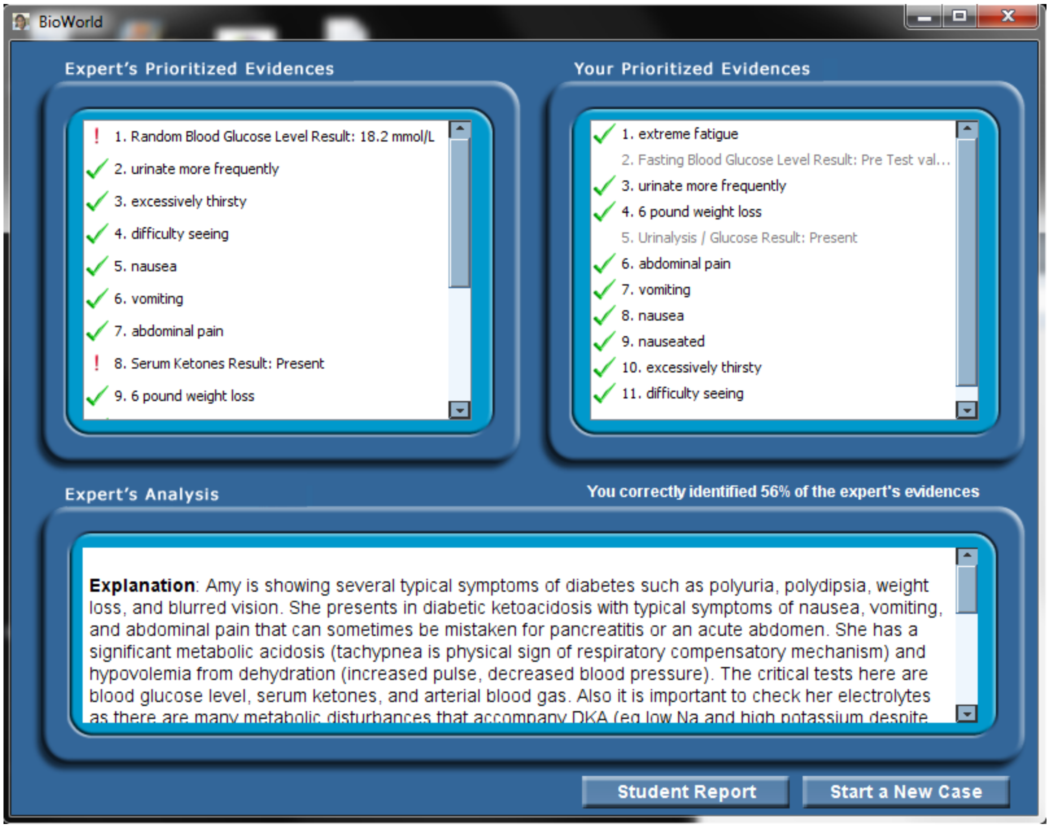
1	2	3	4	5
Deficient knowledge. Needed specific instruction at most steps		Knew all important steps of operation		Demonstrated familiarity with all aspects of operation

OVERALL ON THIS TASK, SHOULD THE CANDIDATE: FAIL PASS

Image from "Testing Technical Skill Via an Innovative 'Bench Station' Examination," by R. Reznick, G. Regehr, H. MacRae, J. Martin, & W. McCulloch, 1997, *The American Journal of Surgery, 173*, 229.

Appendix J

Screen Capture of Expert Solution in BioWorld



Note: red exclamation mark indicates non-match with expert; green checkmark indicates match with expert.

Appendix K
Sample Questions from Deteriorating Patient Simulation Post-test

Note. Correct answers appear in boldface.

In a patient experiencing anaphylactic shock the next priority after removing the allergen and assessing the airway is to administer:

a. Hydrocortisone (Solucortef), 250 mg/kg IV
b. 0.9 % NACL (Normal Saline) 20ml/kg IV
c. Salbutamol (Ventolin), 1 cc in 2cc Normal Saline via ventimask
d. Epinephrine (Adrenalin), 0.5 mg IM
e. Norepinephrine (Levophed), 8-12 mcg/min IV to keep BP>90 systolic

While assessing the ABCD's for a deteriorating patient you note a problem with the patient's "B - Breathing". Your next action should be to:

a. Review the patient's "A – Airway" to ensure it is still clear.
b. Stabilize/treat the "B – Breathing" problem before assessing "C – Circulation".
c. Note the "B – Breathing" problem and assess the patient's "C – Circulation"
d. Note the "B – Breathing" problem and assess the patient's "A, B, C & D" before deciding what to do about it.
e. Prescribe Salbutamol Q1H PRN and Ipatropium Q6H by ventimask, first doses STAT

You are called to see a 70 year-old female because the nurse found her newly unresponsive to painful stimuli. She was admitted to the ward 5 days ago with a right hemiplegia from a left middle cerebral embolic stoke secondary to atrial fibrillation in the setting of prior diabetes and hypertension. On arrival you find the patient unresponsive to painful stimuli and not breathing spontaneously. The next best step is to:

a. Check the airway
b. Check the capillary blood glucose
c. Call a "Code Blue"
d. Check the pulse
e. Give 50 cc of 50% Dextrose IV STAT

You are assessing a patient with Acute Pulmonary Edema and have just administered a first dose of IV furosemide. You note the following on your exam: accessory muscle use, inability to speak in full sentences, O2 sat of 88% on a non-rebreather mask, BP 165/90, HR 100/min regular, no chest pain. While waiting for transfer to ICU for BiPAP you should prescribe:

a. Dobutamine drip at 10 mcg/kg/min IV
b. Propanolol, 40 mg IV q6h first dose STAT
c. Nitrospray 0.4mg sublingual as needed now
d. Digoxin loading dose, 0.75 mg IV now
e. Metalozone 5 mg orally now

12. You are called to see an 84 year-old male because of a BP of 80/60. He was admitted 5 days earlier because of COPD with falls at home complicated by an L1 vertebral body fracture. On arrival you find him somnolent but arousable to voice, moving all limbs symmetrically,

mumbling about someone having given him “a mickie” and with bilateral small pupils. BP 80/60 HR: 50/minute regular RR: 6/minute O2SAT: 89% on room air T: 35 Centigrade rectally. The most important next step at this point is to:

a. Prescribe 50 cc of 50% Dextrose and 100 mg of Thiamine IV
b. Check the ECG
c. Prescribe a 500 cc bolus of 0.9%NACL (Normal Saline)
d. Prescribe 1 mg of Naloxone (Narcan) IV
e. Check an arterial blood gas

9. You are called to see a 59 year-old female diabetic patient who was witnessed having 15 seconds of loss of consciousness with generalized tonic-clonic movements that resolved spontaneously. A capillary blood glucose was undetectable or “low”. On your arrival the patient opens her eyes to painful stimuli, responds by moaning and mumbling “Ouch! That hurts!” and moves all four limbs very well in attempting to punch you. BP: 160/100 HR: 110/min regular O2sat 98% on room air RR: 24/min Temperature: Patient refused. ECG shows sinus tachycardia with biphasic t waves and voltage criteria for left ventricular hypertrophy.
The most important next step is to:

a. Administer 100 grams of liquid glucose sublingual.
b. Administer 5 MG of Librium orally and 100 MG of Thiamine IV.
c. Administer 10 cc of 10% calcium gluconate IV
d. Order STAT serum Na, K, CL, Creatinine & Glucose
e. Administer 50 cc of 50% Dextrose IV

You are called to see an 84 year-old male because of a BP of 80/60. He was admitted 5 days earlier because of COPD with falls at home complicated by an L1 vertebral body fracture. On arrival you find him somnolent but arousable to voice, moving all limbs symmetrically, mumbling about someone having given him “a mickie” and with bilateral small pupils. BP 80/60 HR: 50/minute regular RR: 6/minute O2SAT: 89% on room air T: 35 Centigrade rectally. The most important next step at this point is to:

a. Prescribe 50 cc of 50% Dextrose and 100 mg of Thiamine IV
b. Check the ECG
c. Prescribe a 500 cc bolus of 0.9%NACL (Normal Saline)
d. Prescribe 1 mg of Naloxone (Narcan) IV
e. Check an arterial blood gas

Chapter 5

Final Discussion

Emotions are pervasive throughout medical training. For instance, recalling moments of triumph or disappointment during learning, testing, and professional practice inherently entails an emotional component. Similarly, the innate drive towards pleasant rather than unpleasant events reflects emotional motivations. These emotional experiences are likely to impact actions and performance in the moment, as well as long-term outcomes, such as identity formation and career decisions. Even negative emotions can hold a social and evolutionary utility by providing opportunities to learn from experience and prevent high-risk or problematic behaviors.

Despite its omnipresence, research on emotions in medical education is a relatively new development. Traditionally, the medical literature has placed emotions on the periphery of empirical inquiry. In recent years, there has been a surge of attention and calls for empirical investigations into the role of emotions in the health professions. This dissertation responds to these calls by treating emotions as the focal point of this program of research. Existing affect-related research in this field has typically focused on stress responses as a product of learning environments. However, there is a need to examine a more comprehensive range of emotional experiences to understand how emotions relate to learning and performance. Moreover, to move the field forward, there is a need to ground these investigations in theoretical frameworks appropriate for medical education settings to maximize implications for educational practice through a more systematic approach to studying relations between emotions, other human factors, and facets of the learning environment. These considerations are particularly important within the domain of medicine, given the potential consequences that emotional states can have have on immediate learning and performance, as well as the longer-term affects on employee burnout and quality of patient care. The work presented in this dissertation addresses these gaps in the literature and offers unique contributions to advance the field.

Contributions

In the current dissertation, I drew on theoretical frameworks and measures from education, psychology, affective sciences, and health professions to guide the research design (i.e., constructs, hypotheses, analyses). This was informed by a critical review of the literature (Chapter 2) in which I mapped out key areas and research questions that would advance emotion research in this field. As concluded in Chapter 2, there was a need to test control-value theory within medical domains given its relevance for achievement settings and well-established utility in other education environments. Another conclusion of Chapter 2 was the need to re-examine assumptions underlying the commonly conceptualized structure and classification of emotions. Finally, the review offered considerations for analyses, such as the use of quantitative and qualitative research paradigms, expert/novice comparisons, and change over time from an event-based perspective. This review contributes to the literature in its own right by: (1) identifying limitations in emotion research and theory; (2) providing scaffolds for researchers to embark on investigations of emotion in medical education; and (3) synthesizing and evaluating a complex body of literature for application to the domain of medicine. These contributions also laid the groundwork for the empirical studies of this dissertation.

The two empirical manuscripts (Chapter 3 and 4) contribute uniquely to the literature by offering new emotional states (sampling from four relevant taxonomies of social, epistemic, basic, achievement emotions), new antecedent constructs (control/value appraisals), and a novel and important contribution of examining emotions in relation to performance within authentic medical learning environments. Rather than designing a contrived learning task designed to manipulate emotions, I examined emotions *in situ*, as they spontaneously occur in their natural environment. In the first empirical manuscript (Chapter 3), I surveyed the domain of cardiac surgery to identify a relevant and commonly conducted procedure within the operating room that bears on patient outcomes and surgical skill development. In the second empirical manuscript (Chapter 4), I expanded the sample to include authentic computer-based and simulation environments for diagnostic reasoning and emergency medicine intervention. In both cases, the research contributed to the literature in unique ways by: (1) focusing on state rather than trait emotions to examine emotions as they occur in practice rather than habitual behaviors or aptitudes; (2) using quantitative and qualitative research traditions to triangulate findings, assess validity, and identify areas underspecified by current theories to explore in future work (e.g., context-specific antecedents and regulation strategies for emotions); and (3) employing mixed methodologies involving different data channels (e.g., questionnaires, interviews, electrodermal activity, observational data). In addition, Chapter 3 provides a rare comparison of processes activated during a real case and a simulation. In doing so, it highlights differences in emotional experiences and identifies a key area deserving attention in simulation research; namely, the *emotional fidelity* of simulations. In other words, attending to the extent to which simulations (or other learning activities designed to model real-world problem-solving) approximate the emotional experience of real-life situations may play a critical role in preparing trainees to manage their emotions in clinical practice. This also has implications more broadly for the design of learning activities that are intended to prepare students for high-stakes achievement tasks. This study also developed a new model for simulation that can be used to train and assess performance among surgical trainees prior to conducting an important surgical procedure (vein harvesting) for real cases in the operating room. Chapter 4 uniquely added to the range of analytic approaches by employing both exploratory and confirmatory statistical analyses, as well as drawing on person-centered (within) and group-centered (between) modes of inquiry to examine alignment between emotion measures. The antecedents identified in interviews from both empirical manuscripts can be used as a framework to enhance emotional fidelity, modify instructional design, and guide the development of emotion scales. Figure 13 provides a sample of distal antecedents (focused on features of the learning environments) for researchers to consider in these pursuits. Figure 14 provides an overview of the purpose, methods, and contributions of each manuscript.

Conclusions

On the basis of these two empirical studies, I have made several key conclusions. Specifically, the findings from both studies provide strong support for control-value theory of achievement emotions in medical education. The findings illuminate important relations between appraisals, emotions, and performance, which are consistent with control-value theory in that higher perceived control leads to higher performance and is mediated by the role of emotions. In particular, negative activating emotions, which are most intense among novices, predict lower performance, and are negatively related to control. This suggests that efforts aimed at enhancing control (e.g., through clear instruction, team support, feedback[55]) can help to improve performance by lowering the intensity of negative emotions. Importantly, this research demonstrates that emotion regulation—specifically reappraisal strategies—can help to improve performance. On the other hand, positive activating emotions, appeared to be more related to motivational facets of learning (satisfaction, confidence), rather than performance. Perhaps by teasing apart the influence of discrete positive emotions, more direct relations to performance

[55] Consideration of the timing and nature of implementation for each element is critical. For instance, feedback that focuses extensively on peer comparisons and is perceived as unfair could lead to feelings of frustration. On the other hand, feedback that is timely, specific, and provides learner-focused recommendations for improvement can help to improve skills (Ibrahim, Duffy, Lachapelle, under review).

may be revealed. Figure 15 and 16 provide an overview of possible methods to promote more adaptive emotions from a time sequence perspective (Figure 15) and from an agent/learning environment design perspective (Figure 16).

This research also demonstrates that self-report scales (particularly those using single-item adjectives) can be used as a valid measure of emotion experiences, given their close alignment with other channels of emotion measurement, such as retrospective interviews. It also provides insight into *how* participants interpret and respond to these types of scales, and in doing so, contributes to the development and validation of emotion measurement. Finally, the findings from both manuscripts also indicate the potential value of *novelty* as a third appraisal not currently included within control-value theory. In both manuscripts, epistemic emotions (e.g., surprise, confusion, curiosity) were linked to the novelty and fidelity of the task and may be particularly important for complex or dynamic learning tasks.

Future Directions

The individual manuscripts succinctly outline limitations and future directions (for an overview see Figure 17). Here I elaborate on remaining questions that require further empirical scrutiny as they were beyond the scope of this dissertation; these inform the program of research described next. There are three key themes to moving this research forward: (1) assessing the validity of emotion measures by examining consistency in individuals' perceptions of the meaning of different emotion words; calibration between physiological, observational, and self-report measures as affect-detection technologies are improved; and structure of latent factors and underlying principle components linking discrete emotional states; (2) developing and testing educational interventions designed to promote perceived control and effective emotion regulation, including the use of simulations as a method for emotional preparedness; (3) expanding the scope of this work by examining emotion and regulation across a broader range of medical environments (e.g., classrooms, simulations, online/social media, clinical practice) competencies (e.g., teamwork, patient communication, procedural skills, clinical reasoning, decision-making) and agents (e.g., individual vs teams); (4) using more highly controlled experimental designs to identify mechanisms responsible for the link between emotion and performance (e.g., cognitive load, attentional processes, physiological arousal, judgment, decisions-making); (5) examining emotional states in relation to emotional aptitudes or traits (e.g., emotional intelligence) that can be used for recruitment, selection and training in medicine;

(6) examining how emotions relate not only to performance in the moment but also to learning gains, resilience, and skill transfer over a longer trajectory.

These areas will allow future work to build on the research reported in this dissertation by addressing several unanswered questions, including: At which milestones during learning and training are less intense emotions ideal? When should higher-stress simulations be introduced? Does training for emotion regulation improve performance? Do simulations increase emotional preparedness for clinical practice? How do emotions compare to other human factors in terms of accounting for variability in performance? How do emotions ultimately bear on patient outcomes and patient safety concerns? What type of regulation strategies is most commonly used? Which strategies are most effective? How does monitoring and selection of emotion regulation strategies affect their usefulness? What are the mechanisms underlying the link between emotions and performance in medical education for subsets of competencies? How do emotions in the moment relate to emotional aptitudes or traits? What types of measures are most valid and realistic for assessing emotions in medical education? Do people experience the same emotion differently or is there stability in the meaning of emotions across individuals?

Addressing these complex questions will require integration of theories, methodologies and research paradigms from diverse disciplines (e.g., education, psychology, affective sciences, health sciences). Ultimately, pursuing these lines of inquiry will help to move the field forward by developing an enriched understanding of emotions that no longer places this construct at the periphery of medical training but as a critical psychological process that plays a role in the development of competencies and mastery of a domain. This has bearing not only on medical education, but also more broadly, provides a unique context in which to understand and examine the role of emotions on human performance and learning.

Closing Comments

In fulfilling the requirements of this doctoral dissertation, this research program has led to important theoretical advances in the application of emotion theories to medical education, methodological advances in the use of multi-method research designs to measure emotions and learning variables, and practical advances for the design of medical education environments, including simulation training. Collectively, the manuscripts presented here investigated how emotions relate to important facets of the learning experience and performance across authentic learning environments. By focusing data collection on clinical and simulation environments, this research provides important, but rarely acquired, insight into the nature and role of emotions in medical education for real-world training. In doing so, this research lays the groundwork for educational interventions designed to promote adaptive emotions and effective regulation. It also contributes to a broader discussion about competencies, expertise, and curriculum for medical professionals by including emotions in the conversation.

Distal antecedents (learning environment)	Potential emotions *high* level of antecedent	Potential emotions *low* level of antecedent
Task novelty	Curiosity, surprise, confusion	Relaxation, relief, boredom
Time constraints	Anxiety, frustration	Enjoyment, relaxation
Supportive social dynamics	Gratitude, compassion, hope, pride	Anger, frustration, hopelessness, disgust
High-stakes outcome	Anxiety, fear, shame, guilt, pride	Enjoyment, relaxation, boredom
Evaluative environment	Anxiety, shame, pride, relief, frustration	Enjoyment, relaxation
Formative feedback	Curiosity, hope, pride	Confusion, hopelessness, frustration
Task complexity	Curiosity, surprise, confusion, frustration	Relaxation, relief, boredom

Figure 13. Antecedents to consider for emotional fidelity and scale development

Note. Antecedents can produce positive *or* negative emotions of varying intensity depending on learner characteristics (e.g., prior experience, strategies) and outcomes (success or failure).

Literature Review: Theories and Measures of Emotions

- **Purpose:** Conduct a critical synthesis and review of emotions theories and methods for medical education
- **Methods:** Literature review of psychology, education, affective sciences, health sciences education on topic of emotions in learning
- **Contributions:** Identified advantages and limitations of various theories and methods used to conceptualize and measure emotions; identified critical areas to examine role and nature of emotions in medical education; generated recommendations for research design to advance understanding

Study 1: Emotions in Surgical Training

- **Purpose:** Use a situated approach to examine relations between appraisals, emotions, and performance across conditions (real case v. simulation)
- **Methods:** Questionnaires, interviews, observations, electrodermal activity
- **Sample:** Medical trainees and experts in cardiac surgery operating room
- **Contributions:** Identified expert-novice differences in emotions, antecedents and regulation of emotions; demonstrated significant relations between control, emotions, and regulation consistent with Control-value Theory; highlighted importance of psychological fidelity of simulation; developed simulation model for training and assessment

Study 2: Emotions across Medical Education Environments

- **Purpose:** Test generalizability of findings from surgical study; examine validity of self-report scale and nature of emotions
- **Methods**: Questionnaires, interviews, log files, observations
- **Sample**: Medical trainees in 3 environments: operating room, computer-based learning environment, emergency care simulation
- **Contributions**: Identified how participants respond and interpret to self-report scales; established alignment between scales and retrospective interview accounts; demonstrated further support for control-value theory

Figure 14. Overview of purpose, methods, and contributions of each manuscript.

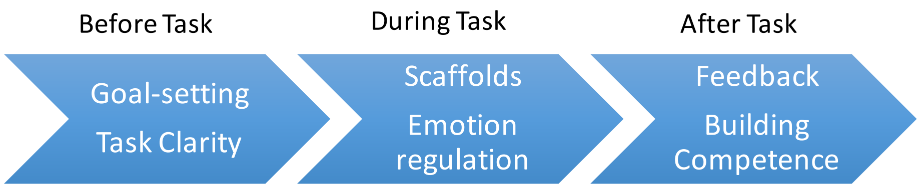


Figure 15. Sample strategies to promote more adaptive emotions over time.
Note. This process is considered recrusive rather than strictly linear

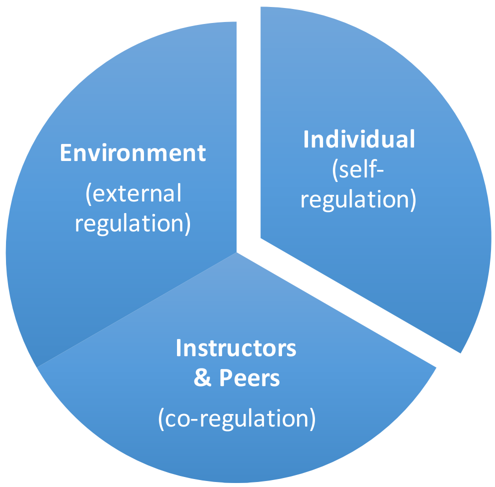


Figure 16. Sources of regulation to promote more adaptive emotions.
Note. In this research (Chapter 3) the focus was regulation efforts emanating from the individual.

Key Limitations	Future Directions
Sample size limited for certain statistical analyses (e.g., confirmatory factor analysis)	Expand sample size, continue to test across a range of medical environments, tasks, and competencies
Limited attention toward differences between environments and causal links	Greater standardization/control over measures, tasks, prior experience, use of experimental design
Limited attention toward differences between discrete emotions	Conduct analyses focusing on discrete emotions in future work
Some potentially relevant discrete emotions excluded from measure	Create long-form questionnaire with more exhaustive range of emotions
Focused on value for activity or task rather than value/importance of outcome, use of single-item scales	Examine the multi-faceted nature of control and value appraisals using multi-item scales
Focused on performance and self-efficacy rather than other facets of learning	Examine learning gains (pre-post), emotional preparedness, skill transfer
Measured emotional experiences and regulation at the individual level rather than social level	Examine social emotions/co-regulation that are shared in team environments
Tasks were relatively short in duration compared to semester-long or repeated-practice longitudinal data	Examine change in emotion over longer periods of time for more time-intensive tasks/clinical rotations

Figure 17. Summary of limitations and future directions of empirical work.

Bibliography

Affectiva. (2014). Q U Manual. Available at: http://qsensor-support.affectiva.com/customer/portal/articles/1539375-q-user-manual. Accessed November 2015.

Aggarwal, R., Grantcharov, T., Moorthy, K., Milland, T., & Darzi, A. (2008). Toward feasible, valid, and reliable video-based assessments of technical surgical skills in the operating room. *Annals of Surgery, 247,* 372–379.

Aggarwal, R., Mytton, O. T., Derbrew, M., Hananel, D., Heydenburg, M., Issenberg, B., … Reznick, R. (2010). Training and simulation for patient safety. *Quality and Safety in Health Care*, *19*(Suppl 2), i34–i43.

Agha, R. A., Fowler, A. J., & Sevdalis, N. (2015). The role of non-technical skills in surgery. *Annals of Medicine and Surgery*, *4*(4), 422–427.

Aldao, A., & Nolen-Hoeksema, S. (2010). Specificity of cognitive emotion regulation strategies: A transdiagnostic examination. *Behaviour Research and Therapy, 48,* 974–983.

Aldao, A., & Nolen-Hoeksema, S. (2012). When are adaptive strategies most predictive of psychopathology? *Journal of Abnormal Psychology, 121,* 276–281.

Allsop, J., & Gray, R. (2014). Flying under pressure: Effects of anxiety on attention and gaze behavior in aviation. *Journal of Applied Research in Memory and Cognition*, *3*(2), 63–71.

American College of Cardiology and American Heart Association Task Force. (2004). Guideline update for coronary artery bypass graft surgery. *Journal of the American College of Cardiology, 44,* 1146–1154.

Angie, A. D., Connelly, S., Waples, E. P., & Kligyte, V. (2011). The influence of discrete emotions on judgement and decision-making: A meta-analytic review. *Cognition and Emotion*, *25*(8), 1393–1422.

Antiel, R. M., Van Arendonk, K. J., Reed, D. A., Terhune, K. P., Tarpley, J. L., Porterfield, J. R., … Farley, D. R. (2012). Surgical training, duty-hour restrictions, and implications for meeting the Accreditation Council for Graduate Medical Education core competencies: views of surgical interns compared with program directors. *Archives of Surgery*, *147*(6), 536–541.

Arora, S., Sevdalis, N., Aggarwal, R., Sirimanna, P., Darzi, A., & Kneebone, R. (2010). Stress impairs psychomotor performance in novice laparoscopic surgeons. *Surgical Endoscopy and Other Interventional Techniques, 24,* 2588–2593.

Arora, S., Sevdalis, N., Nestel, D., Tierney, T., Woloshynowych, M., Kneebone, R. (2009). Managing intraoperative stress: What do surgeons want from a crisis training program? *The American Journal of Surgery, 197*, 537–543.

Arora, S., Sevdalis, N., Nestel, D., Woloshynowych, M., Darzi, A., & Kneebone, R. (2010). The impact of stress on surgical performance: A systematic review of the literature. *Surgery*, *147*(3), 318–330.

Arora, S., Sevdalis, N., Suliman, I., Athanasiou, T., Kneebone, R., & Darzi, A. (2009). What makes a competent surgeon? Experts' and trainees' perceptions of the roles of a surgeon. *The American Journal of Surgery*, *198*(5), 726–732.

Artino, A. J., & Durning, S. J. (2011). It's time to explore the role of emotion in medical students' learning. *Academic Medicine, 86*, 275–276.

Artino, A. R. (2013). When I say…emotion in medical education. *Medical Education, 47,* 1062–1063.

Artino, A. R., Holmboe, E. S. & Durning, S. J. (2012). Control-value theory: Using achievement emotions to improve understanding of motivation, learning, and performance in medical education: *Medical Teacher 34*, 148–160.

Artino, A. R., La Rochelle, J. S., & Durning, S. J. (2010). Second year medical students' motivational beliefs, emotions, and achievement. *Medical Education*, *44*, 1203–1212.

Artino, A. R., & Naismith, L. M. (2015). "But how do you really feel?" Measuring emotions in medical education research. *Medical Education*, *49*(2), 140–142.

Aucar, J. A., Groch, N. R., Troxel, S. A., & Eubanks, S. W. (2005). A review of surgical simulation with attention to validation methodology. *Surgical Laparoscopy, Endoscopy & Percutaneous Techniques*, *15*(2), 82–89.

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

Azenberg, Y., & Picard, R. (2014). Feel: A system for frequent event and electrodermal activity labeling. *IEEE Journal of Biomedical and Health Informatics, 18,* 266–277.

Azevedo, R., Harley, J., Trevors, G., Duffy, M., Feyzi-Behnagh, R., Bouchet, F., & Landis, R. S. (2013). Using trace data to examine the complex roles of cognitive, metacognitive, and emotional self-regulatory processes during learning with multi-agent systems. In R. Azevedo & V. Aleven (Eds.), *International handbook of metacognition and learning technologies* (pp. 427–449). NY: Springer.

Babineau, T. J., Becker, J., Gibbons, G., Sentovich, S., Hess, D., Robertson, S., & Stone, M . (2004). The "cost" of operative training for surgical residents. *Archives of Surgery*, *139*(4), 366–370.

Baker, G., Norton, P., Flintoff, V., Blais, R., Brown, A., Cox, J., … Tamblyn, R. (2004). The Canadian adverse events study: The incidence of adverse events among hospital patients in Canada. *Canadian Medical Association Journal, 170,* 1678–1686.

Baker, C. J., Sinha, R., & Sullivan, M. E. (2012). Development of a cardiac surgery simulation curriculum: From needs assessment results to practical implementation. *The Journal of Thoracic and Cardiovascular Surgery*, *144*(1), 7–16.

Bandura, A. (2006). Guide for constructing self-efficacy scales. In F. Pajares & T. Urdan (Eds.), *Self-efficacy beliefs of adolescents* (Vol. 5, pp. 307–337). Greenwich, CT: Information Age Publishing.

Bandyopadhyay, D., Pammi, V. C., & Srinivasan, N. (2013). Role of affect in decision making. *Progress in brain research*, *202*, 37–53.

Barrett, L. F. (2009). The future of psychology: Connecting mind to brain. *Perspectives in Psychological Science*, *4*, 326–339.

Bebko, G. M., Franconeri, S. L., Ochsner, K. N., & Chiao, J. Y. (2011). Look before you regulate: Differential perceptual strategies underlying expressive suppression and cognitive reappraisal. *Emotion*, *11*(4), 732–742.

Becker, E. S., Goetz, T., Morger, V., & Ranellucci, R. (2014). The importance of teachers' emotions and instructional behavior for their students' emotions—An experience sampling analysis. *Teaching and Teacher Education, 43*, 15–26.

Beilock, S. L. (2008). Math performance in stressful situations. *Current Directions in Psychological Science, 17,* 339–343.

Ben-Eliyahu, A., & Linnenbrink, L. (2013). Extending self-regulated to include self-regulated emotion strategies. *Motivation and Emotion, 37*, 558–573.

Berguer, R., Smith, W. D., & Chung, Y. H. (2001). Performing laparoscopic surgery is significantly more stressful for the surgeon than open surgery. *Surgical Endoscopy, 15,* *1204*–1207.

Blanchette, I., & Richards, A. (2010). The influence of affect on higher level cognition: A review of research on interpretation, judgement, decision making and reasoning. *Cognition and Emotion*, *24*(4), 561–595.

Blum, M. G., Powers, T. W., & Sundaresan, S. (2004). Bronchoscopy simulator effectively prepares junior residents to competently perform basic clinical bronchoscopy. *The Annals of Thoracic Surgery*, *78*(1), 287–291.

Boekaerts, M. (2011). Emotions, emotion regulation, and self-regulation of learning. In B. Zimmerman & D. Schunk (Eds.). *Handbook of self-regulation of learning and performance* (pp. 408–425). New York: Routledge.

Bond, W. F., Deitrick, L. M., Arnold, D.C., Kostenbader, M., Barr, G. C., Kimmel, S. R., & Worrilow, C. C. (2004). Using simulation to instruct emergency medicine residents in cognitive forcing strategies. *Academic Medicine*, 79, 438–446.

Bridges, M., & Diamond, D. L. (1999). The financial impact of teaching surgical residents in the operating room. *The American Journal of Surgery*, *177*(1), 28–32.

Brydges, R., Carnahan, H., Rose, D., Rose L., & Dubrowski, A. (2010). Coordinating progressive levels of simulation fidelity to maximize educational benefit. *Academic Medicine, 85*(5), 806–812.

Brydges, R., Manzone, J., Shanks, D., Halata, R., Hamstra, S. J., Zendejas, B., & Cook, D. A. (2015). Self-regulated learning in simulation-based training: A systematic review and meta-analysis. *Medical Education, 49*, 368–378.

Buckley, C. E., Kavanagh, D. O., Traynor, O., & Neary, P. C. (2014). Is the skillset obtained in surgical simulation transferable to the operating theatre? *The American Journal of Surgery*, *207*(1), 146–157.

Butler, D. L., & Brydges, R. (2013). Learning in the health professions: what does self-regulation have to do with it? *Medical Education, 47,* 1057–1059.

Buss, D. M. (2014). Comment: Evolutionary criteria for considering an emotion “basic”: Jealousy as an illustration. *Emotion Review, 4,* 313–331.

Calvo, R. A., & D’Mello, S. (2012). Frontiers of affect-aware learning technologies. IEEE *Intelligent Systems, 27*, 86–89.

Cannon, W. B. (1927). The James-Lange theory of emotions: A critical examination and an alternative theory. *The American Journal of Psychology*, *39,* 106–124.

Campos, J. J., Walle, E. A., Dahl, A., & Main, A. (2011). Reconceptualizing emotion regulation. *Emotion Review, 3*, 26–35.

Cheung, R. Y., & Au, T. K. (2011). Nursing students' anxiety and clinical performance. *Journal of Nursing Education, 50*, 286–289.

Chew, B. H., Zain, A. M., & Hassan, F. (2013). Emotional intelligence and academic performance in first and final year medical students: A cross-sectional study. *BMC Medical Education*, *13*, 1–10.

Chi, M. T. H. (2006). Two approaches to the study of experts' characteristics. In K. A. Ericsson, N. Charness, P. Feltovich, & R. Hoffman (Eds.), *The Cambridge handbook of expertise and expert performance* (pp. 121-30), New York: Cambridge University Press.

Clarke, S., Horeczko, T., Cotton, D., & Bair, A. (2014). Heart rate, anxiety and performance of residents during a simulated critical clinical encounter: A pilot study. *BMC Medical Education, 14*, 1–8.

Cook, D. A., Brydges, R., Ginsburg, S., & Hatala, R. (2015). A contemporary approach to validity arguments: a practical guide to Kane's framework. *Medical Education, 49,* 560–575.

Cook, D. A., Zendejas, B., Hamstra, S. J., Hatala, R., & Brydges, R. (2014). What counts as validity evidence? Examples and prevalence in a systematic review of simulation-based assessment. *Advances in Health Sciences Education, 19,* 233–250.

Coombs, S. A., Cauraugh, J. H., & Jannell, C. M. (2007). Emotional State and Initiating Cue Alter Central and Peripheral Motor Processes. *Emotion, 7*(2), 275284.

Craig, S. D., D'Mello, C. S., Witherspoon A., & Graesser, A. (2008). Emote-aloud during learning with AutoTutor: Applying the Facial Action Coding System to cognitive-affective states during learning. *Cognition and Emotion, 22*, 777–788.

Croskerry, P. (2003). Cognitive forcing strategies in clinical decision-making. *Annals of Emergency Medicine, 41*, 110–120.

Croskerry, P., Abbass, A. A., & Wu, A. W. (2008). How doctors feel: Affective issues in patients' safety. *The Lancet, 372,* 1205–1206.

Csikszentmihalyi, M., & Larson, R. (1987). Validity and reliability of the experience sampling method. *Journal of Nervous and Mental Diseases, 9*, 526–536.

Cuff, P. A., & Vanselow, N. (Eds). (2004). *Improving medical education: Enhancing the behavioral and social science content of medical school curricula*. Washington DC: National Academies Press.

Cutuli, D. (2014). Cognitive reappraisal and expressive suppression strategies role in the emotion regulation: an overview on their modulatory effects and neural correlates. *Frontiers in Systems Neuroscience, 8,* 1–6.

Creedon, P. J., & Hayes, A. F. (2015, May). *Small sample mediation analysis: How far can you push the bootstrap?* Paper presented at the annual conference of the Association for Psychological Science, New York, USA.

Darwin, C. (1872). *The expression of the emotions in man and animals*. London, UK: John Murray.

Dehning, S., Gasperi, S., Krause, D., Meyer, S., Reiß, E., Burger, M., … Siebeck, M. (2013). Emotional and cognitive empathy in first-year medical students. *ISRN Psychiatry*, *2013*, 801530.

de Montbrun, S. L., & MacRae, H. (2012). Simulation in Surgical Education. *Clinics in Colon and Rectal Surgery*, *25*(3), 156–165.

de Vries, E. N., Ramrattan, M. A., Smorenburg, S. M., Gouma, D. J., & Boermeester, M. A. (2008). The incidence and nature of in-hospital adverse events: A systematic review. *Quality and Safety in Health Care, 3,* 216–223.

de Vries, E. N., Prins, H. A., Crolla, R. M. P. H., den Outer, A. J., van Andel, G., van Helden, S. H., … Boermeester, M. A. (2010). Effect of a comprehensive surgical safety system on patient outcomes. *The New England Journal of Medicine*, *363*(20), 1928–1937.

Dimick, J. B., Chen, S. L., Taheri, P. A., Henderson, W. G., Khuri, S. F., & Campbell, D. A. (2004). Hospital costs associated with surgical complications: A report from the private-sector National Surgical Quality Improvement Program. *Journal of the American College of Surgeons*, *199*(4), 531–537.

D'Mello, S. (2013). A selective meta-analysis on the relative incidence of discrete affective states during learning with technology. *Journal of Educational Psychology, 105,* 1082–1099.

D'Mello, S. K., Dale, R. A., & Graesser, A. C. (2012). Disequilibrium in the mind, disharmony in the body. *Cognition and Emotion, 26*, 362–374.

D'Mello, S. K., & Graesser, A. C. (2009). Automatic detection of learners' emotions from gross body language. *Applied Artificial Intelligence, 23*(2), 123–150.

D'Mello, S. K., & Graesser, A. C. (2010). Multimodal semi-automated affect detection from conversational cues, gross body language, and facial features. *User Modeling and User-adapted Interaction, 20* (2), 147–187.

D'Mello, S. K., & Graesser, A. C. (2012). Dynamics of affective states during complex learning. *Learning and Instruction, 22,* 145–157.

D'Mello, S., Lehman, B., Pekrun, R., & Graesser, A. (2012). Confusion can be beneficial for learning. *Learning and Instruction, 29,* 153–170.

Doty, D. B., & Doty, J. R. (2012). Coronary artery bypass graft. In D. B Doty & J. R. Doty (Eds.), *Cardiac surgery: Operative technique (2nd Ed.)* (pp. 393–431). Philadelphia, PA: Elsivier.

Duffy, M. C., Azevedo, R., Sun, N., Griscom, S., Stead, V., Crelinsten, L., ... Lachapelle, K. (2015a). Team regulation in a simulated medical emergency: An in-depth analysis of cognitive, metacognitive, and affective processes. *Instructional Science, 43*, 401–426.

Duffy, M. C., Lajoie, S., Jarrell, A., Pekrun, R., Azevedo, R., & Lachapelle, K. (2015b, April). *Emotions in medical education: Developing and testing a scale of emotions across medical learning environments*. Paper presented at the annual meeting of the American Educational Research Association, Chicago, IL.

Duffy, M. C., Lajoie, S., & Lachapelle, K. (2016). Measuring emotions in medical education: Methodological and technological advances within authentic medical learning environments. In S. Bridges, L. K. Chan, & C. E. Hmelo-Silver (Eds.) (pp. 181–213). *Educational technologies in contextual learning: Research in health professions education*. Switzerland: Springer.

Duffy, M. C., Lajoie, S., Pekrun, R., Ibrahim, M., & Lachapelle, K. (2015c, March). *Simulation versus operating room: Examining relations between emotional states, learning processes, and technical skills performance across training environments*. Poster presented at the annual meeting of the consortium of American College of Surgeons Accredited Education Institutes, Chicago, IL.

Dunn, B. D., Evans, D., Makarova, D., White, J., & Clark, L. (2012). Gut feelings and the reaction to perceived inequity: The interplay between bodily responses, regulation, and perception shapes the rejection of unfair offers on the ultimatum game. *Cognitive, Affective, & Behavioral Neuroscience, 12*, 149–429.

Dunnington, G. L. (1996). The art of mentoring. *American Journal of Surgery*, *171*(6), 604–607.

Dyrbye, L. N., West, C. P., Satele, D., Boone, S., Tan, L., Sloan, J., & Shanfelt, T. D. (2014). Burnout among U.S. medical students, residents, and early career physicians relative to the general U.S. population. *Academic Medicine, 89,* 443–451.

Eccles, J. S., Adler, T. F., Futterman, R., Goff, S. B., Kaczala, C. M., Meece, J. L., & Midgley, C. (1983). Expectancies, values, and academic behaviors. In J. T. Spence (Ed.), *Achievement and achievement motivation* (pp. 75–146). San Francisco, CA: W. H. Freeman.

Efklides, A. (2011). Interactions of metacognition with motivation and affect in self-regulated learning: The MASRL model. *Educational Psychologist, 46*(1), 6–25.

Eisenberg, N., Fabes, R. A., Guthrie, I. K., & Reiser, M. (2000). Dispositional emotionality and regulation: Their role in predicting quality of social functioning. *Journal of Personality and Social Psychology, 78*, 136–157.

Ekman, P. (1972). Universal and cultural differences in facial expressions of emotions. In J. K. Cole (Ed.), *Nebraska symposium on motivation, 1971* (pp. 207–283). Lincoln, NE: University of Nebraska Press.

Ekman, P. (1992). An argument for basic emotions. *Cognition and Emotion*, *6*, 169–200.

Ekman, P. (1999). Basic emotions. In T. Dalgleish, & M. Power (Eds.), *Handbook of Cognition and Emotion* (pp. 45–60). Sussex, UK: John Wiley & Sons.

Ekman, P., & Cordar, D. (2011). What is meant by calling emotion basic. *Emotion Review, 3,* 364–370.

Ekman, P., Friesen, W. V., & Hager, J. C. (2002). *Facial action coding system* [E-book]. Salt Lake City, UT: Research Nexus.

Ellsworth, P. C. (2013). Appraisal theory: Old and new questions. *Emotion Review*, *5*, 125–131.

Emotion [Def. 2c]. (n.d.). *Merriam-Webster Online*. Retrieved September 24, 2014, from http://www.merriam-webster.com/dictionary/emotion.

Ericsson, K., & Simon, H. (1993). *Protocol analysis: Verbal reports as data* (2nd ed.). Boston: MIT Press.

Eva, K. W. (2009). Broadening the debate about quality in medical education research. *Medical Education, 43*, 294–296.

Farkouh, M. E., Domanski, M., Sleeper, L. A., Siami, R. S., Dangas, G., Mack, M., … Fuster, V. (2012). Strategies for multivessel revascularization in patients with diabetes. *New England Journal of Medicine, 367*, 2375–2384.

Fawver, B., Amano, S., Hass, C. J., & Janelle, C. M. (2012). Active control of approach-oriented posture is influenced by emotional reactions. *Emotion*, *12*(6), 1350–1361.

Fedor, S., & Picard, R. (2014). Ambulatory EDA: Comparisons of bilateral forearm and calf locations. Poster presented at the 54th Annual Meeting of the Society of Psychophysiological Research. Atlanta, USA.

Fisher, A. H., & van Kleef, G. A. (2010). Where have all the people gone? A plea for including social interaction in emotion research. *Emotion Review, 2,* 208–211.

Flin, R., Yule, S., Paterson-Brown, S., Maran, N., Rowley, D., & Youngson, G. (2007). Teaching surgeons about non-technical skills. *The Surgeon*, *5*(2), 86–89.

Fontaine, J. J. R., & Scherer, K. R. (2013). The global meaning structure of the emotion domain: Investigating the complementarity of multiple perspectives on meaning. In J. J. R. Fontaine, K. R. Scherer, & C. Soriano (Eds.), *Components of emotional meaning*. *A sourcebook* (pp. 106–125). Oxford: Oxford University Press.

Fontaine, J. R. J., Scherer, K. R., Soriano, C. (2013). *Components of emotional meaning: A sourcebook*. Oxford: Oxford University Press.

Fraser, K., Huffman, J., Ma, I., Sobczak, M., McIlwrick, J., Wright, B., & McLaughlin, K. (2014). The emotional and cognitive impact of unexpected simulated patient death: A randomized controlled trial. *Chest, 145*, 958–963.

Fraser, K., Ma, I., Teteris, E., Baxter, H., Wright, B., & McLaughlin, K. (2012). Emotion, cognitive load and learning outcomes during simulation training. *Medical Education, 46*, 1055–1062.

Frijda, N. H. (1986). *The emotions*. New York, NY: Cambridge University Press.

Frijda, N. H. (1987). Emotion, cognitive structure, and action tendency. *Cognition and Emotion*, *1*, 115–143.

Frijda, N. H. (2007). *The laws of emotion*. Mahwah, N.J.: L. Erlbaum Associates.

Fritz, M. S., & MacKinnon, D. P. (2007). Required sample size to detect the mediated effect. *Psychological Science, 18*, 233–239.

Gaba, D. M. (2004). The future vision of simulation in health care. *Quality and Safety in Health Care*, *13*(suppl 1), i2–i10.

Gardner, D. G., Cummings, L. L., Dunham, R. B., & Pierce, J. L. (1998). Single-item versus multiple-item measurement scales: An empirical comparison. *Educational and Psychological Measurement*, *58*(6), 898–915.

Gauthier, G., & Lajoie, S.P. (2014). Do expert clinical teachers have a shared understanding of what constitutes a competent reasoning performance in case-based teaching? *Instructional Science, 42,* 579-594.

Gawande, A. A., Zinner, M. J., Studdert, D. M., & Brennan, T. A. (2003). Analysis of errors reported by surgeons at three teaching hospitals. *Surgery*, *133*(6), 614–621.

Gélat, T., Coudrat, L., & Le Pellec, A. (2011). Gait initiation is affected during emotional conflict. *Neuroscience Letters*, *497*(1), 64–67.

Global Burden of Disease 2013 Mortality and Causes of Death, Collaborators (2015). Global, regional, and national age-sex specific all-cause and cause-specific mortality for 240 causes of death, 1990-2013: A systematic analysis for the Global Burden of Disease Study 2013. *The Lancet*, *385*, 117–171.

Goetz, T., Bieg, M., Lüdtke, O., Pekrun, R., & Hall, N. C. (2013). Do girls really experience more anxiety in mathematics? *Psychological Science*, *24*(10), 2079–2087.

Goetz, T., Frenzel, A., Hall, N. C., Nett, U. E., Pekrun, R., & Lipnevich, A. A. (2014). Types of boredom: An experience sampling approach. *Motivation & Emotion, 38,* 401-419.

Goetz, T., Frenzel, A. C., Pekrun, R., Hall, N. C., Lüdtke, O. (2007). Between- and within-domain relations of students' academic emotions. *Journal of Educational Psychology, 99*, 715–ß733.

Goetz, T., Pekrun, R., & Hall, N.,& Haag, L. (2006). Academic emotions from a social-cognitive perspective: Antecedents and domain specificity of students' affect in the context of Latin instruction. *The British Psychological Society, 76*, 289–308.

Gogol, K., Brunner, M., Goetz, T., Martin, R., Ugen, S., Keller, U., … Preckel, F. (2014). "My questionnaire is too long!" The assessments of motivational-affective constructs with three-item and single-item measures. *Contemporary Educational Psychology*, *39*(3), 188–205.

Graesser, A., D'Mello, S., & Strain, A. (2014). Emotions in advanced learning technologies. In R. Pekrun & L. Linnenbrink-Garcia (Eds.), *International handbook of emotions in education* (pp. 289–310). New York: Routledge

Greenberg, C. C., Regenbogen, S. E., Studdert, D. M., Lipsitz, S. R., Rogers, S. O., Zinner, M. J., & Gawande, A. A. (2007). Patterns of Communication Breakdowns Resulting in Injury to Surgical Patients. *Journal of the American College of Surgeons*, *204*(4), 533–540.

Greene, J. A., Dellinger, K., Binbasaran Tuysuzoglu, B., & Costa, L. (2013). A two-tiered approach to analyzing self-regulated learning process data to inform the design of hypermedia learning environments. In R. Azevedo, & V. Aleven (Eds.), *International handbook of metacognition and learning technologies* (pp. 117–128). New York: Springer.

Greene, J. A., Yu, S. B., & Copeland, D. Z. (2014). Measuring critical components of digital literacy and their relationships with learning. *Computers & Education, 76,* 55–69.

Gross, J. J. (1998a). Antecedent- and response-focused emotion regulation: Divergent consequences for experience, expression, and physiology. *Journal of Personality and Social Psychology, 74,* 224–237.

Gross, J. J. (1998b). The emerging field of emotion regulation: An integrative review. *Review of General Psychology, 2,* 271–299.

Gross, J. J. (2001). Emotion regulation in adulthood: Timing is everything. *Current Directions in Psychological Science, 10,* 214–219.

Gross, J. J. (2015). The extended process model of emotion regulation: Elaborations, applications, and future directions. *Psychological Inquiry, 26,* 130–137.

Gross, J. J., & Barrett, L. F. (2011) Generation and emotion regulation: One or two depends on your point of view. *Emotion Review, 3,* 8–16.

Gross, J. J., & John, O. P. (2003). Individual differences in two emotion regulation processes: Implications for affect, relationships and well-being. *Journal of Personality and Social Psychology, 85,* 348–362.

Gross, J. J., Sheppes, G., & Urry, H. L. (2011). Emotion generation and emotion: A distinction we should make (carefully), *Cognition and Emotion, 25*, 765–781.

Gross, J. J., & Thompson, R. A. (2007). Emotion regulation: Conceptual foundations. In J. J. Gross (Ed.), *Handbook of emotion regulation* (pp. 3–24). New York, NY: Guilford Press.

Gulati, R., Rihal, C. S., & Gersh, B. J. (2009). The SYNTAX Trial A Perspective. *Circulation: Cardiovascular Interventions*, *2*, 463–467.

Hailikari, T., Katajavuori, N., & Lindblom-Ylanne, S. (2008). The relevance of prior knowledge in learning and instructional design. *American Journal of Pharmaceutical Education, 72,* 113–121.

Hamdorf, J. M., & Hall, J. C. (2000). Acquiring surgical skills. *The British Journal of Surgery*, *87*, 28–37.

Hamstra, S. J., Byrdges, R., Hatala, R., Zendejas, B., & Cook, D. A. (2014). Reconsidering fidelity in simulation-based training. *Academic Medicine, 89,* 387–392.

Harvey, A., Bandiera, G., Nathens, A. B., & LeBlanc, V. R. (2012). Impact of stress on resident performance in simulated trauma scenarios. *Journal of Trauma Acute Care Surgery 72,* 497–503.

Hassan, I., Weyers, P., Maschuw, K., Dick, B., Gerdes, B., Rothmund, M., & Zielke, A. (2006). Negative stress-coping strategies among novices in surgery correlate with poor virtual laparoscopic performance. *British Journal of Surgery, 93*, 1554–1559.

Hayes, A. F., & Preacher, K. J. (2014). Statistical mediation analysis with a multicategorical independent variable. *British Journal of Mathematical and Statistical Psychology, 67,* 451-470.

Haynes, A. B., Weiser, T. G., Berry, W. R., Lipsitz, S. R., Breizat, A. H. S., Dellinger, E. P., … Gawande, A. A. (2009). A surgical safety checklist to reduce morbidity and mortality in a global population. *New England Journal of Medicine*, *360*(5), 491–499.

Hektner, J. M., Schmidt, J. A., & Csikszentmihalyi, M. (2007). Experience sampling method: Measuring the quality of everyday life. Thousand Oaks, CA: Sage Publications, Inc.

Hess, C. N., Lopes, R. D., Gibson, C. M., Hager, R., Wojdyla, D. M., Englum, B. R., … Alexander, J. H. (2014). Saphenous vein graft failure after coronary artery bypass surgery: Insights from PREVENT IV. *Circulation*, 130, 1445–1451.

Hull, L., Arora, S., Aggarwal, R., Darzi, A., Vincent, C., & Sevdalis, N. (2012). The impact of nontechnical skills on technical performance in surgery: A systematic review. *Journal of the American College of Surgeons*, *214*, 214–230.

Hunziker, S., Laschinger, L., Portmann-Schwarz, S., Semmer, N.K., Tschan, F., & Marsch, S. (2011). *Perceived stress and team performance during a simulated resuscitation. Intensive Care Medicine, 37,* 1473–1479.

Hunziker, S., Pagani, S., Fasleer, K., Tschan, F., Semmer, N. K., & Marsch, S. (2013). Impact of a stress coping strategy on perceived stress levels and performance during a simulated cardiopulmonary resuscitation. *BMC Emergency Medicine, 13*, 8–17.

Hunziker, S., Semmer, N.K., Tschan, F., Schuetz, P., Mueller, B., & Marsch, S. (2012). Dynamics and association of different acute stress markers with performance during a simulated resuscitation. *Resuscitation, 83*, 572–578.

Ibrahim, M., Duffy, M. C., & Lachapelle, K. (2016, August). *The use of the surgical skills rubric improves technical skill acquisition and retention*. Poster to be presented at the Association for Medical Education in Europe Annual Conference. Barcelona, Spain.

James, W., & Lange, C. G. (1922). *The emotions*. Baltimore: Williams & Wilkins Co.

Immordino-Yang, M. H., & Christodoulou, J. A. (2014). Neuroscientific contributions to understanding and measuring emotions in educational contexts. In R. Pekrun & L. Linnenbrink-Garcia (Eds.), *International handbook of emotions in education* (pp. 607-624). New York: Routledge

Issenberg, S., McGaghie, W. C., Hart, I. R., Mayer, J. W., Felner, J. M., Petrusa, E. R., … Ewy, G. A. (1999). Simulation technology for health care professional skills training and assessment. *Journal of the American Medical Association*, *282*(9), 861–866.

Izard, C. E. (1993). Four systems for emotion activation: Cognitive and noncognitive processes. *Psychological Review*, *100*, 68–90.

Izard, C. E. (2007). Basic emotions, natural kinds, emotion schemas, and a new paradigm. *Perspectives on Psychological Science*, *2*, 260–280.

Izard, C. E. (2010). The many meanings/aspects of emotion: Definitions, functions, activation, and regulation, *Emotion Review, 2,* 363–370.

Izard, C. E. (2011). Forms and functions of emotions: Matters of emotion–cognition interactions. *Emotion Review*, *3*, 371–378.

James, W. (1890). *The principles of psychology*. New York, NY: Dover.

Järvenoja, H., & Järvelä, S. (2009). Emotion control in collaborative learning situations: Do students regulate emotions evoked by social challenges. *British Journal of Educational Psychology, 79*, 463–481.

Järvenoja, H., Volet, S., Järvelä, S. (2013). Regulation of emotions in socially challenging learning situations: An instrument to measure the adaptive and social nature of the regulation process. *International Journal of Experimental Educational Psychology, 33*, 31–58.

Kairys, J. C., McGuire, K., Crawford, A. G., & Yeo, C. J. (2008). Cumulative operative experience is decreasing during general surgery residency: A worrisome trend for surgical trainees? *Journal of the American College of Surgeons*, *206*(5), 804–811.

Kaplan, S., Bradley, J. C., Luchman, J. N., & Haynes, D. (2009). On the role of positive and negative affectivity in job performance: A meta-analytic investigation. *Journal of Applied Psychology*, *94*, 162–176.

Karabenick, S. A., Woolley, M. E., Friedel, J. M., Ammon, B. V., Blazevski, J., Bonney, C. R., … Musu, L. (2007). Cognitive processing of self-report items in educational research: Do they think what we mean? *Educational Psychologist*, *42*(3), 139–151.

Kasman, D. L., Fryer-Edwards, K., & Braddock, C. H. (2003). Educating for professionalism: Trainees' emotional experiences on IM and pediatrics inpatient ward. *Academic Medicine, 78,* 730–741.

Kavussanu, M., Dewar, A. J., & Boardley, I. D. (2014). Achievement goals and emotions in athletes: The mediating role of challenge and threat appraisals. *Motivation and Emotion*, *38*(4), 589–599.

Keitel, A., Ringleb, M., Schwartges, I., Weik, U., Picker, O., Stockhorst, U., & Deinzer, R. (2011). Endocrine and psychological stress responses in a simulated emergency situation. *Psychoneuroendocrinology, 36*, 98–108.

Keltner, D., & Buswell, B. N. (1997). Embarrassment: Its distinct form and appeasement functions. *Psychological Bulletin, 122,* 250–270.

Kenton, K. (2006). How to Teach and Evaluate Learners in the Operating Room. *Obstetrics and Gynecology Clinics*, *33*, 325–332.

Kerr, B., & O'Leary, J. P. (1999). The training of the surgeon: Dr. Halsted's greatest legacy. *The American Surgeon*, *65*, 1101–1102.

Kneebone, R. (2005). Evaluating clinical simulations for learning procedural skills: a theory-based approach. *Academic Medicine: Journal of the Association of American Medical Colleges*, *80*(6), 549–553.

Kohn, K. T., Corrigan, J. M., & Donaldson, M. S. (2000). *To err is human: Building a safer health system*. Washington, DC: National Academy Press.

Kotsis, S. V., & Chung, K. C. (2013). Application of the "see one, do one, teach one" concept in surgical training. *Plastic and Reconstructive Surgery*, *131*(5), 1194–1201.

Kozlowski, S. W. J. & DeShon, R. P. (2004). A psychological fidelity approach to simulation-based training: Theory, research, and principles. In E. Salas, L. R. Elliott, S. G. Schflett, & M. D. Coovert (Eds.), *Scaled Worlds: Development, validation, and applications* (pp. 75–99). Burlington, VT: Ashgate Publishing.

Kreibig, S. (2010). Autonomic nervous system activity in emotion: A review. *Biological Psychology, 84,* 394–421.

Kreibig, S. D., & Gendolla, G. H. E. (2014). Autonomic nervous system measurement of emotion in education and achievement settings. In R. Pekrun, & L. Linnenbrink-Garcia (Eds.), *International handbook of emotions in education* (pp. 625–642). New York, NY: Routledge.

Kring, A. M., & Sloan, D. M. (2007). The facial expression coding system (FACES): Development, validation, and utility. *Psychological Assessment, 19,* 210–224.

Lajoie, S. P. (2009). Developing professional expertise with a cognitive apprenticeship model: Examples from Avionics and Medicine. In K. A. Ericsson (Ed.). *Development of Professional Expertise: Toward Measurement of Expert Performance and Design of Optimal Learning Environments* (pp. 61–83). New York: NY. Cambridge University Press.

Lajoie, S., Poitras, E., Naismith, L., Gauthier, G., Summerside, C., & Kazemitabar, M. (2013). Modelling domain-specific self-regulatory activities in clinical reasoning. In H. C. Lane, K. Yacef, & J. Mostow (Eds.), *International artificial intelligence and education proceedings* (pp. 632–635). Amsterdam: IOS Press.

Larsen, J. T., & McGraw, P. (2011). Further evidence for mixed emotions. *Journal of Personality and Social Psychology, 100*, 1095–1110.

Larsen, C. R., Oestergaard, J., Ottesen, B. S., & Soerensen, J. L. (2012). The efficacy of virtual reality simulation training in laparoscopy: a systematic review of randomized trials. *Acta Obstetricia et Gynecologica Scandinavica*, *91*(9), 1015–1028.

Lazarus, R. S. (1966). *Psychological stress and the coping process*. New York, NY: McGraw-Hill.

Lazarus, R. S. (1991). *Emotion and adaptation*. New York, NY: Oxford University Press.

Lazarus R. S., & Folkman, S. (1984). *Stress, appraisal, and coping*. New York: Springer.

LeBlanc, V. R. (2009). The effects of acute stress on performance: implications for health professions education. *Academic Medicine, 84,* 25–33.

LeBlanc, V. R., & Bandiera. G. W. (2007). The effects of examination stress on the performance of emergency medicine residents. *Medical Education, 41,* 556-564.

LeBlanc, V. R., MacDonald, R. D., McArthur, B., King, K., & Lepine, T. (2005). Paramedic performance in calculating drug dosages following stressful scenarios in a human patient simulator. *Prehospital Emergency Care, 9,* 439–444.

LeBlanc, V. R., Manser, T., Weinger, M. B., Musson, D., Kutzin, J., & Howard, S. K. (2011). The study of factors affecting human and systems performance in healthcare using simulation. *Simulation in Healthcare, 6,* 24–29.

LeBlanc, V. R., McConnell, M. M., & Monteiro, S. D. (2014). Predictable chaos: a review of the effects of emotions on attention, memory and decision-making. *Advances in Health Sciences Education*, *20,* 265–282.

LeBlanc, V. R., Regehr, C., Tavares, W., Scott, A.K., Macdonald, R., & King, K. (2012**).** The impact of stress on paramedic performance during simulated critical events. *Prehospital and Disaster Medicine, 27*, 369–374.

LeBlanc, V. R., Woodrow, S. I., Sidhu, R., & Dubrowski, A. (2008). Examination stress leads to improvements on fundamental technical skills for surgery. *American Journal of Surgery, 196,* 114–119.

LeDoux, J. E. (2014). Comment: What's basic about the brain mechanisms of emotions? *Emotion Review 6*, 318–320.

Leonard, M., Graham, S. & Bonacum, D. (2004). The human factor: The critical importance of effective teamwork and communication in providing safe care. *Quality and Safety in Health Care,* 13, 85–90.

Levine, E. L., Xu, X., Yang, L., Ispas, D., Pitariu, H. D., Bian, R., … Musat, S. (2011). Cross-national explorations of the impact of affect at work using the state-trait emotion measure: A coordinated series of studies in the United States, China, and Romania. *Human Performance, 24,* 405–442.

Lindeman, B. M., Sacks, B. C., Hirose, K., & Lipsett, P. A. (2014). Duty hours and perceived competence in surgery: are interns ready? *Journal of Surgical Research*, *190*(1), 16–21.

Lindström, B. R., & Bohlin, G. (2011). Emotion processing facilitates working memory performance. *Cognition and Emotion*, *25*(7), 1196–1204.

Lingard, L., Espin, S., Whyte, S., Regehr, G., Baker, G. R., Reznick, R., … Grober, E. (2004). Communication failures in the operating room: an observational classification of recurrent types and effects. *Quality & Safety in Health Care*, *13*(5), 330–334.

Lingard, L., Regehr, G., Orser, B., Reznick, R., Baker, G. R. Doran, D., Espin, S., Bohnen, J., & Whyte, S. (2008). Evaluation of a preoperative checklist and team briefing among surgeons, nurses, and anesthesiologists to reduce failures in communication. *Archives of Surgery 143*, 12–17.

Loderer, K., Gentsch, K., Duffy, M. C., Pekrun, R., & Scherer, K., Meier, E., Fontaine J. R., & Soriano, C. (under review). Profiling achievement emotions: A control-value approach using the achievement emotions CoreGRID. *Motivation and Emotion*.

MacRae, H., Regehr, G., Leadbetter, W., & Reznick, R. K. (2000). A comprehensive examination for senior surgical residents. *The American Journal of Surgery, 179*, 190–193.

Manser, T., & Foster, S. (2011). Effective handover communication: An overview of research and improvement efforts. *Safety in Anaesthesia 25,* 181–91.

Marteau, T. M., & Bekker, H. (1992). The development of a six-item short-form of the state scale of the Spielberger State-Trait Anxiety Inventory (STAI). *British Journal of Clinical Psychology, 31,* 301–306.

Martin, J. A., Regehr, G., Reznick, R., Macrae, H., Murnaghan, J., Hutchison, C. & Brown, M. (1997). Objective structured assessment of technical skill (OSATS) for surgical residents. *British Journal of Surgery, 84*, 273–278.

Matejka, M., Kazzer, P., Seehausen, M., Bajbouj, M., Klann-Delius, G., Menninghaus, W., Jacobs, A. M., Heekeren, H. R., & Prehn, K. (2013). Talking about emotion: Prosody and skin conductance indicate emotion regulation. *Frontiers in Psychology, 4,* 1–11.

Matsumoto, D., & Hwang, H. S. (2011). Evidence for training the ability to read microexpressions of emotion. *Motivation and Emotion*, *35*(2), 181–191.

Mattarella-Micke, A., Mateo, J., Kozak, M. N., Foster, K., & Beilock, S. L. (2011). Choke or thrive? The relation between salivary cortisol and math performance depends on individual differences in working memory and math-anxiety. *Emotion, 11,* 1000–1005.

Mauss, I. B., Cook, C. K., & Gross, J. J. (2007). Automatic emotion regulation during anger provocation. *Journal of Experimental Social Psychology, 43,* 698–711.

Mauss, I. B., & Robinson, M. D. (2009). Measures of emotions: A review. *Cognition and Emotion, 23*, 209–237.

McConnell, M. M. & Eva, K. W. (2012). The role of emotion in the learning and transfer of clinical skills and knowledge. *Academic Medicine, 87,* 1316–1322.

McIntire, S. A. & Miller, L. A. (2007). *Foundations of psychological testing* (2nd ed.). Thousand Oaks, CA: Sage.

McWilliams, L. A., Cox, B. J., & Enns, M. W. (2003). Use of coping inventory for stressful situations in a clinically depressed sample: Factor structure, personality correlates, and prediction of distress. *Journal of Clinical Psychology, 59*, 423–437.

Meinhardt, J., & Pekrun, R. (2003). Attentional resource allocation to emotional events: An ERP study. *Cognition and Emotion, 17*, 477–500.

Mesquita, B. (2010). Emoting: A contextualized process. In B. Mesquita, L. F. Barrett & E. Smith (Eds.), *The mind in context* (pp. 83–104). New York, NY: Guilford.

Mesquita, B., & Boiger, M. (2014). Emotions in context: A sociodynamic model of emotions. *Emotion Review*, *6*, 298–302.

Meunier, J., Merckaert, I., Libert, Y., Delvaux, N., Etienne, A.-M., Lienard, A., Bragard, I., Marchal, S., Reynaert, C., Slachmuylder, J. L., & Razavi, D. (2013). The effect of communication skills training on residents' physiological arousal in a breaking bad news simulated task. *Patient Education & Counseling, 93,* 40–47.

Meyers, L. S., Gamst, G., & Guarino, A. J. (2013). Applied multivariate research: Design and interpretation (2nd ed.). Thousand Oaks, CA: Sage.

Mitmasgruber, H., Beck, T. N., Höfer, S., & Schübler, G. (2009). When you don't like what you feel: Experiential avoidance, mindfulness, and meta-emotion in emotion regulation. *Personality and Individual Differences, 46,* 448–453.

Mohr, F. W., Morice, M. C., Kappetein, A P., Feldman, T. E., Ståhle, E., Colombo, A., … Serruys, P. W. (2013). Coronary artery bypass graft surgery versus percutaneous coronary intervention in patients with three-vessel disease and left main coronary disease: 5-year follow-up of the randomised, clinical SYNTAX trial. *The Lancet, 381,* 629–638.

Moors, A. (2014). Flavors of appraisal theories of emotion. *Emotion Review, 6*, 303–307.

Moors, A., Ellsworth, P. C., Scherer, K. R., & Frijda, N. H. (2013). Appraisal theories of emotion: State of the art and future development. *Emotion Review, 5*, 119-124.

Muis, K. R., Duffy, M. C, Trevors, G., & Ranellucci, J., & Foy, M. (2014). What were they thinking? Using cognitive interviewing to examine the validity of self-reported epistemic beliefs. *International Journal of Education, 2,* 17–32.

Müller, M. P., Hansel, M., Fichtner, A., Hardt, F., Weber, S., Kirschbaum, C., Rüder, S., Walcher, F., Koch, T., & Eich, C. (2009). Excellence in performance and stress reduction during two different full scale simulator training courses: A pilot study. *Resuscitation, 80,* 919–924.

Mulligan, K., & Scherer, K. R. (2012). Toward a working definition of emotion. *Emotion Review, 4,* 345-357.

Mundfrom, D. J., Shaw, D. G., & Lu Ke, T. (2005). Minimum sample size recommendations for conducting factor analyses. *International Journal of Testing, 5,* 159-168.

Murphy, F. C., Nimmo-Smith, I., & Lawrence, A. D. (2003). Functional neuroanatomy of emotions: a meta-analysis. *Cognitive Affective & Behavioral Neuroscience, 3,* 207–233.

Nasca, T. J., Day, S. H., Amis, E. S., & ACGME Duty Hour Task Force. (2010). The new recommendations on duty hours from the ACGME Task Force. *The New England Journal of Medicine*, *363*(2), e3.

National Heart, Lung, and Blood Institute. (2012). What is coronary artery bypass grafting? Available at: http://www.nhlbi.nih.gov/health/health-topics/topics/cabg/. Accessed

January, 2015.

Naugle, K. M., Joyner, J., Hass, C. J., & Janelle, C. M. (2010). Emotional influences on locomotor behavior. *Journal of Biomechanics*, *43*(16), 3099–3103.

Naugle, K. M., Coombes, S. A., Cauraugh, J. H., & Janelle, C. M. (2012). Influence of Emotion on the control of low-level force production. *Research Quarterly for Exercise and Sport*, *83*(2), 353–358.

Nesse, R. M. (2014). Comment: A general "theory of emotion" is neither necessary nor possible. *Emotion Review*, *6*, 320–322.

Nesse, R. M., & Ellsworth, P. C. (2009). Evolution, emotions, and emotional disorders. *American Psychologist*, *6*, 129–139.

Nett, U. E., Goetz, T., & Hall, N. C. (2011). Coping with boredom in school: An experience sampling perspective. *Contemporary Educational Psychology, 36*, 49–59.

O'Callaghan, A. (2013). Emotional congruence in learning and health encounters in medicine: Addressing as aspect of the hidden curriculum. *Advances in Health Sciences Education*, *18,* 305–317.

O'Connor, T., Papanikolaou, V., & Keogh, I. (2010). Safe surgery, the human factors approach. *The Surgeon*, *8*(2), 93–95.

O'Hagan J., MacKinnon N. J., Persaud D., & Etchegary, H. (2009). Self-reported medical errors in seven countries: Implications for Canada. *Healthcare Quarterly, 12,* 55–61.

Öhman, A., Hamm, A., & Hugdahl, K. (2000). Cognition and the autonomic nervous system: Orienting, anticipation, and conditioning. In: J. T. Cacioppo, L. G. Tassinary, & G. G. (Eds.), *Handbook of psychophysiology (*Vol. 2, pp. 533–575). New York: Cambridge University Press.

Ortony, A., & Turner, T. J. (1990). What's basic about basic emotions? *Psychological Review*, *97*, 315–331.

Pajares, F. (1996). Self-efficacy beliefs in achievement settings. *Review of Educational Research, 66,* 543578.

Palter, V. N., & Grantcharov, T. P. (2014). Individualized deliberate practice on a virtual reality simulator improves technical performance of surgical novices in the operating room: a randomized controlled trial. *Annals of Surgery*, *259*(3), 443–448.

Panksepp, J., & Watt, D. What is basic about basic emotions? Lasting lessons from affective neuroscience. *Emotion Review, 3,* 387–396.

Pantic, M., & Rothkrantz, L. J. M. (2000). Automatic analysis of facial expressions: The state of the art. *Pattern Analysis and Machine Intelligence, IEEE Transactions on, 22*(12), 1424–1445.

Pekrun, R. (1992). The impact of emotions on learning and achievement: Towards a theory of cognitive/motivational mediators. *Applied Psychology, 41*, 359-376.

Pekrun, R. (2006). The control-value theory of achievement emotions: Assumptions, corollaries, and implications for educational research and practice. *Educational Psychology Review, 18*, 315–341.

Pekrun, R., & Bühner, M. (2014). Self-report measures of academic emotion. In R. Pekrun & L. Linnenbrink-Garcia (Eds.), *International handbook of emotions in education* (pp. 561–579). New York: Routledge

Pekrun, R., Elliot, A. J., & Maier, M. A. (2009). Achievement goals and achievement emotions: Testing a model of their join relations with academic performance. *Journal of Educational Psychology, 101,* 115–135.

Pekrun, R. (2006). The control-value theory of achievement emotions: Assumptions, corollaries, and implications for educational research and practice. *Educational Psychology Review, 18*, 315–341.

Pekrun, R., Frenzel, A. C., Goetz, T., & Perry, R. P. (2007). The control-value theory of achievement emotions: An integrative approach to emotions in education. In P. A. Schutz & R. Pekrun (Eds.), *Emotion in education* (pp. 13–36). San Diego, CA: Academic Press.

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,* 36–48.

Pekrun, R., Goetz, T., & Perry, R. P. (2005). Achievement Emotions Questionnaire (AEQ). User's manual. *Munich, Germany: University of Munich, Department of Psychology*.

Pekrun, R., Goetz, T., Titz, W., & Perry, R. P. (2002). Academic emotions in students' self-regulated learning and achievement: A program of qualitative and quantitative research. *Educational Psychologist*, *37*, 91–105.

Pekrun, R., & Linnenbrink-Garcia, L. (2014a). Introduction to emotions in education. In R. Pekrun & L. Linnenbrink-Garcia (Eds.), *International handbook of emotions in education* (pp. 1–10). New York: Taylor & Francis/Routledge.

Pekrun, R., & Linnenbrink-Garcia, L. (Eds.). (2014b). *International handbook of emotions in education*. New York: Taylor & Francis/Routledge.

Pekrun, R., & Meier, E. (2011). Epistemic Emotion Scales (EES). Unpublished manuscript, Department of Psychology, University of Munich, Munich Germany.

Pekrun, R., Meier, E., Muis, K. R., & Sinatra, G. M. (revise and resubmit). Measuring emotions related to knowledge and knowing: The Epistemic Emotions Scales (ESS). *Cognition and Emotion*.

Pekrun, R., & Perry, R. P. (2014). Control-value theory of achievement emotions. In R. Pekrun & L. Linnenbrink-Garcia (Eds.), *International handbook of emotions in education* (pp. 120–141). New York: Routledge.

Pekrun, R., & Stephens, E. J. (2012). Academic emotions. In K. R. Harris, S. Graham, T. Urdan, S. Graham, J. M. Royer, & M. Zeidner (Eds.), *APA educational psychology handbook* (Vol. 2, pp. 3–31). Washington, DC: American Psychological Association.

Picard, R. W. (2010). Emotion research by the people, for the people. *Emotion Review, 2,* 250–254.

Pintrich, P. R. (2000). The role of goal orientation in self-regulated learning. In M. Boekaerts, P. Pintrich & M. Zeidner (Eds.), *Handbook of self-regulation* (pp. 451–502). San Diego, CA: Academic Press.

Pintrich, P. (2004). A conceptual framework for assessing motivation and self-regulated learning in college students. *Educational Psychology Review, 16,* 385-407.

Piquette, D., Reeves, S., & LeBlanc, V. R. (2009). Stressful intensive care unit medical crises: How individual responses impact on team performance. *Critical Care Medicine, 37*, 1251–1255.

Piquette, D., Tarshis, J., Sinuff, T., Pinto, R., & LeBlanc, V. R. (2014). Impact of acute stress on resident performance during simulated resuscitation episodes: a perspective randomized cross-over study. *Teaching and Learning in Medicine: An International Journal, 26,* 9–16.

Plutchik, R. (1980). A general psychoevolutionary theory of emotion. In R. Plutchik & H. Kellerman (Eds.), *Emotion: Theory, research, and experience: Theories of emotion* (Vol. 1, pp. 3–33). New York: Academic.

Plutchik, R. (2001). The nature of emotions. *American Scientist*, *89*, 344–350.

Poh, M., Swenson, N. C., & Picard, R., W. (2010). A wearable sensor for unobtrusive, long-term assessment of electrodermal activity. *IEEE*, *Transactions on Biomedical Engineering, 57,* 1243–1252.

Posner, J., Russell, J. A., & Peterson, B. S. (2005). The circumplex model of affect: An integrative approach to affective neuroscience, cognitive development, and psychopathology. *Developmental Psychopathology, 17,* 715-734.

Preacher, K. J., & Hayes, A. F. (2008). Contemporary approaches to assessing mediation in communication research. In A. F. Hayes, M. D. Slater, and L. B. Snyder (Eds), *The Sage sourcebook of advanced data analysis methods for communication research* (pp. 13–54). Thousand Oaks, CA: Sage Publications.

Quartz, S. R. (2009). Reason, emotion and decision-making: risk and reward computation with feeling. *Trends in Cognitive Sciences*, *13*(5), 209–215.

Regehr, G. (2010). It's not rocket science: Rethinking our metaphors for research in health professions education. *Medical Education, 44,* 31–39.

Reisenzein, R., Junge, M., Studtmann, M., & Huber, O. (2014). Observational approaches to the measurement of emotions. In R. Pekrun & L. Linnenbrink-Garcia (Eds.), *International handbook of emotions in education* (pp. 580–606). New York: Routledge.

Reznick, R. K., & MacRae, H. (2006). Teaching surgical skills—Changes in the wind. *The New England Journal of Medicine, 355,* 2664–2669.

Reznick, R., Regehr, G., MacRae, H., Martin, J., McCulloch, W. (1997). Testing technical skill via an innovative "bench station" examination. *American Journal of Surgery*. *173*(3), 226–230.

Rogers Jr, S. O., Gawande, A. A., Kwaan, M., Puopolo, A. L., Yoon, C., Brennan, T. A., & Studdert, D. M. (2006). Analysis of surgical errors in closed malpractice claims at 4 liability insurers. *Surgery*, *140*(1), 25–33.

Russell, J. A. (2003). Core affect and the psychological construction of emotion. *Psychological Review*, *110*, 145–172.

Russell, J. A., & Barrett, L. F. (1999). Core affect, prototypical emotional episodes, and other things called emotion: Dissecting the elephant. *Journal of Personality and Social Psychology*, *7*, 805–819.

Sachdeva, A. K. (2002). Acquisition and maintenance of surgical competence. *Seminars in Vascular Surgery*, *15*(3), 182–190.

Sachdeva, A. K., Buyske, J., Dunnington, G. L., Sanfey, H. A., Mellinger, J. D., Scott, D. J., … Burns, K. J. (2011). A new paradigm for surgical procedural training for surgical procedural training: *Current Problems in Surgery*, *48*(12), 854–968.

Satterfield, J. M., & Hughes, E. (2007). Emotion skills training for medical students: A systematic review. *Medical Education, 41,* 935–941.

Scarantino, A., & Griffiths, P. (2011). Don't give up on basic emotions. *Emotion Review, 3,* 444–454.

Schachter, S., & Singer, J. E. (1962). Cognitive, social and physiological determinants of emotional state. *Psychological Review*, *69*, 379–399.

Scherer, K. R. (1984). Emotion as a multicomponent process: A model and some cross-cultural data. In P. Shaver (Ed.), *Review of personality and social psychology* (Vol. 5, pp. 37–63). Beverly Hills, CA: Sage.

Scherer, K. (2005). What are emotions? And how can they be measured? *Social Science Information, 44*, 695-729.

Scherer, K. R. (2009). The dynamic architecture of emotion: Evidence for the component process model. *Cognition and Emotion, 23,* 1307-1351.

Scherer, K. R. (2013). The nature and dynamics of relevance and valence appraisals: Theoretical advances and recent evidence. *Emotion Review*, *5*, 150–162.

Scherer, K. R., Shuman, V., Fontaine, J. R. J., & Soriano, C. (2013). The GRID meets the wheel: Assessing emotional feeling via self-report. In J. R. J. Fontaine, K. R. Scherer, & C. Soriano (Eds.), *Components of emotional meaning: A sourcebook* (pp. 281-298). Oxford, United Kingdom: Oxford University Press.

Sedlack, R. E., & Kolars, J. C. (2004). Computer simulator training enhances the competency of gastroenterology fellows at colonoscopy: Results of a pilot study. *The American Journal of Gastroenterology*, *99*(1), 33–37.

Seymour, N. E., Gallagher, A. G., Roman, S. A., O'Brien, M. K., Bansal, V. K., Andersen, D.

K., & Satava, R. M. (2002). Virtual reality training improves operating room performance. *Annals of Surgery*, *236*(4), 458–464.

Shapiro, J. (2011). Does medical education promote professional alexithymia? A call for attending to the emotions of patients and self in medical training. *Academic Medicine, 86,* 326–332.

Sharma, N., & Gedeon, T. (2012). Objective measures, sensors and computational techniques for stress recognition and classification: A survey. *Computer Methods and Programs in Biomedicine, 103,* 1287–1301.

Sharma, B., Mishra, A., Aggarwal, R., & Grantcharov, T. P. (2011). Non-technical skills assessment in surgery. *Surgical Oncology*, *20*(3), 169–177.

Shouhed, D., Gewertz, B., Wiegmann, D., & Catchpole, K. (2012). Integrating human factors research and surgery: A review. *Archives of Surgery*, *147*(12), 1141–1146.

Shuman, V., & Scherer, K. R. (2014). Concepts and structure of emotions. In R. Pekrun & L. Linnenbrink-Garcia (Eds.), *International handbook of emotions in education* (pp. 13-35). New York: Routledge.

Silvia, P. J. (2005). What is interesting? Exploring the appraisal structure of interest. *Emotion*, *5*, 89–102.

Silvia, P. J. (2009). Looking past pleasure: Anger, confusion, disgust, pride, surprise, and other unusual aesthetic emotions. *Psychology of Aesthetic, Creativity, and the Arts, 3,* 48–51.

Silvia, P. J. (2010). Confusion and interest: The role of knowledge emotions in aesthetic experience. *Psychology of Aesthetics, Creativity, and the Arts, 4,* 75–80.

Smith, C., A., & Ellsworth, P. C. (1985). Patterns of cognitive appraisal in emotion. *Journal of Personality and Social Psychology, 4*, 813–838.

Solomon, R. C. (2003). *Not passion's slave: Emotions and choice*. New York, NY: Oxford University Press.

Spielberger, C. D. (1989). *State-Trait Anxiety Inventory: Bibliography* (2nd ed.). Palo Alto, CA: Consulting Psychologists Press.

Spielberger C. D., Gorsuch R., Lushene R., Vagg P., & Jacobs G. (1983). *Manual for the State-Trait Anxiety Inventory*. Palo Alto, CA: Consulting Psychologists Press.

Stahel, P. F., Sabel, A.L., Victoroff, M. S., Varnell, J., Boygle, D. J., Clarke, T. J., … Mehler, P. S. (2010). Wrong-site and wrong-patient procedures in the universal protocol era:

Analysis of a prospective database of physician self-reported occurrences. *Archives of Surgery*, *145*(10), 978–984.

Starcke, K., & Brand, M. (2012). Decision making under stress: A selective review. *Neuroscience & Biobehavioral Reviews*, *36*(4), 1228–1248.

Stins, J. F., & Beek, P. J. (2011). Organization of voluntary stepping in response to emotion-inducing pictures. *Gait & Posture*, *34*(2), 164–168.

Stuart, E. A. (2010). Matching methods for causal inference: A review and a look forward. *Statistical Science, 25*(1), 1–21.

Sturm, L. P., Windsor, J. A., Cosman, P. H., Cregan, P., Hewett, P. J., & Maddern, G. J. (2008). A systematic review of skills transfer after surgical simulation training. *Annals of Surgery*, *248*(2), 166–179.

Sullivan, G. M. (2011). Getting off the "gold standard": Randomized controlled trials and education research. *Journal of Graduate Medical Education, 3,* 285-289.

Swick, H. M. (2000). Toward a normative definition of medical professionalism. *Academic Medicine, 75*, 612–616.

Tamir, M. (2011). The maturing field of emotion regulation. *Emotion Review, 3*, 3-7.

Tendulkar, A. P., Victorino, G. P., Chong, T. J., Bullard, M. K., Liu, T. H., & Harken, A. H. (2005). Quantification of surgical resident stress "on call". *Journal of American College of Surgeons, 201*, 560–564.

Tracy, J. (2014a). An evolutionary approach to understanding distinct emotions. *Emotion Review*, *6*, 308–312.

Tracy, J. (2014b). Author reply: Incompatible conclusions or different levels of analysis? *Emotion Review*, *6*, 330–331.

Tracy, J. L., & Randles, D. (2011). Four models of basic emotions: A review of Ekman and Cordaro, Izard, Levenson, and Panksepp and Watt. *Emotion Review, 3,* 397-405.

Tracy, J. L., & Robins, R. W. (2004). Show your pride: Evidence for a discrete emotion expression. *Psychological Science, 15,* 194–197.

Tracy, J. L., & Robins, R. W. (2007). The prototypical pride expression: Development of a nonverbal coding scheme. *Emotion, 7*, 789-801.

Tsuda, S., Scott, D., Doyle, J., & Jones, D. B. (2009). Surgical skills training and simulation. *Current Problems in Surgery*, *46*(4), 271–370.

Tugade, M. M., & Fredrickson, B. L. (2007). Regulation of positive emotions: Emotion regulation strategies that promote resilience. *Journal of Happiness Studies, 8,* 311–333.

Turner, J. C., & Trucano, M. (2014). Measuring situated emotion. In R. Pekrun & L. Linnenbrink-Garcia (Eds.), *International handbook of emotions in education (*pp. 643–658). New York: Routledge.

Turner, J. C., & Trucano, M. (2014) Measuring situated emotion. In R. Pekrun & L. Linnenbrink-Garcia (Eds.), *International Handbook of Emotions in Education*. New York: Routledge.

Undre, S., Koutantji, M., Sevdalis, N., Gautama, S., Selvapatt, N., Williams, S., … Vincent, C. (2007). Multidisciplinary crisis simulations: the way forward for training surgical teams. *World Journal of Surgery, 31,* 1843–1853.

van Dulmen, S., Tromp, F., Grosfeld, F., ten Cate, O., & Bensing, J. (2007). The impact of assessing simulated bad news consultations on medical students stress response and communication performance. *Psychoneuroendocrinology, 32*, 943–95.

Van Gog, T., & Jarodzka, H. (2013). Eye tracking as a tool to study and enhance cognitive and metacognitive processes in computer-based learning environments. In R. Azevedo & V. Aleven (Eds.), *International handbook of metacognition and learning technologies* (pp. 143–156). New York, NY: Springer.

Vincent, C., Neale, G., & Woloshynowych, M. (2001). Adverse events in British hospitals: preliminary retrospective record review. *British Medical Journal*, *322*(7285), 517–519.

Walter, A. J. (2006). Surgical education for the twenty-first century: beyond the apprentice model. *Obstetrics and Gynecology Clinics of North America*, *33*(2), 233–236.

Watson, D., & Clark, L. A. (1999). *The PANAS-X: Manual for the positive and negative affect schedule-expanded form*. University of Iowa.

Watson, D., Clark, L. A., & Tellegen, A. (1988). Development and validation of brief measures of positive and negative affect: The PANAS scales. *Journal of Personality and Social Psychology*, *54*, 1063–1070.

Watson, D., & Naragon-Gainey, K. (2010). On the specificity of positive emotional dysfunction in psychopathology: Evidence from the mood and anxiety disorders and schizophrenia/schizotypy. *Clinical Psychology Review*, *30*(7), 839–848.

Webb, T. L., Miles, E., & Sheeran, P. (2012). Dealing with feeling: A meta-analysis of the effectiveness of strategies derived from the process model of emotion regulation. *Psychological Bulletin, 138,* 775–808.

Weiner, B. (2007). Examining emotional diversity on the classroom: An attribution theorist considers the moral emotions. In P. A. Shutz & R. Pekrun (Eds.), *Emotion in education* (pp. 73–88). San Diego, CA: Academic Press.

Wetzel, C. M., Black, S. A., Hanna, G. B., Athanasiou, T., Kneebone, R. L., Nestel, D., Wolfe, J. H. N., & Woloshynowych, M. (2010). The effects of stress and coping on surgical performance during simulations. *Annals of Surgery, 251*, 171–176.

Wetzel, C. M., Kneebone, R. L., Woloshynowych. M., Nestel, D., Moorthy, K., Kidd, J., & Darzi, A. (2006). The effects of stress on surgical performance, *American Journal of Surgery*, *191,* 5–10.

Wigfield, A., & Eccles, J. S. (2000). Expectancy-value theory of achievement motivation. *Contemporary Educational Psychology 25,* 68–81.

Williams, R. G., & Dunnington, G. L. (2004). Accreditation Council for Graduate Medical Education core competencies initiative: The road to implementation in the surgical specialties. *The Surgical Clinics of North America*, *84*(6), 1621–1646.

Winne, P. H. (2001). Self-regulated learning viewed from models of information processing. In B. J., Zimmerman & D. H. Schunk (Eds.), *Self-regulated learning and academic achievement: Theoretical perspectives* (2nd Ed.) (pp. 153–189). Mahwah, NJ: Erlbaum.

Winne, P., & Hadwin, A. (1998). Studying as self-regulated learning. In D. Hacker, J. Dunlosky, & A. Graesser (Eds.), *Metacognition in educational theory and practice* (pp. 227-304). Mahwah, NJ: Erlbaum.

Winne, P., & Hadwin, A. (2008). The weave of motivation and self-regulated learning. In D. Schunk & B. Zimmerman (Eds.), *Motivation and self-regulated learning: Theory, research, and applications* (pp. 297– 314). Mahwah, NJ: Erlbaum.

Wiseman, J. & Labelle, M. (2015, March). *Early recognition of and response to acute reterioration (ERRAD): A theory-based method for teaching medical undergraduates acute care thinking*. Presented at the Canadian Conference on Medical Education.

Wiseman, J., & Snell, L. (2008). The deteriorating patient: A realistic but “low-tech” simulation of emergency decision-making. *The Clinical Teacher, 5*, 93-97.

Wolf, O. T. (2009). Stress and memory in humans: Twelve years of progress? *Brain Research*, *1293*, 142–154.

World Health Organization. (2014). The top 10 causes of death. World health statistics 2012. Available at: http://www.who.int/mediacentre/factsheets/fs310/en/. Accessed October 2015.

Yamamoto, A., Hara, T., Kikuchi, K., & Fujiwara, T. (1999). Intraoperative stress experienced by surgeons and assistants. *Ophthalmic Surgery and Lasers, 30*, 27–30.

Yiend, J. (2010). The effects of emotion on attention: A review of attentional processing of emotional information. *Cognition and Emotion*, *24*(1), 3–47.

Yik, M., Russell, J. A., & Steiger, J. H. (2011). A 12-point circumplex structure of core affect. *Emotion, 11,* 705–731.

Young, J. Q., Van Merrienboer, J., Durning, S., & ten Cate, O. (2014). Cognitive load theory: Implications for medical education: AMEE Guide No. 86, *Medical Teacher, 36*(5), 371–384.

Yu, R. (2015). Choking under pressure: The neuropsychological mechanism of incentive-induced performance decrements. *Frontiers in Behavioral Neuroscience, 9*, 1–8.

Yule, S., Flin, R., Paterson-Brown, S., & Maran, N. (2006a). Non-technical skills for surgeons in the operating room: A review of the literature. *Surgery*, *139*(2), 140–149.

Yule, S., Flin, R., Paterson-Brown, S., Maran, N., & Rowley, D. (2006b). Development of a rating system for surgeons' non-technical skills. *Medical Education*, *40*(11), 1098–1104.

Zeidner, M. (2014). Anxiety in education. In R. Pekrun & L. Linnenbrink-Garcia (Eds.), *International handbook of emotions in education* (pp. 265-288). New York: Routledge.

Zenati, M. A., Shroyer, A. L., Collins, J. F., Hattler, B., Ota, T., Almassi, G. H., … Sonel, A. F. (2011). Impact of endoscopic versus open saphenous vein harvest technique on late coronary artery bypass grafting patient outcomes in the ROOBY (Randomized On/Off Bypass) Trial. *The Journal of Thoracic and Cardiovascular Surgery*, *141*(2), 338–344.

Zeng, Z., Pantic, M., Roisman, G. I., & Huang, T.S. (2009). A survey of affect recognition methods: Audio, visual, and spontaneous expressions. *IEEE Transactions on Pattern Analysis and Machine Learning, 31*, 39–58.

Zentner, M., Grandjean, D., & Scherer, K. (2008). Emotions evoked by the sound of music: Characterization, classification, and measurement. *Emotion*, *8*, 494–521.

Zimmerman, B. (2000). Attaining self-regulation: A social cognitive perspective. In M. Boekaerts, P. Pintrich, & M. Zeidner (Eds.), *Handbook of self-regulation* (pp. 13–39). San Diego, CA: Academic Press.

Zimmerman, B. (2011). Motivational sources and outcomes of self-regualted learning and performance. In B. J. Zimmerman & D. H. Schunk (Eds.), *Handbook of self-regulation of learning and performance* (pp. 49–64). New York: Routledge.