

Quantitative metrics and psychometric scales in the visual art and medical education literature: a narrative review

John David Ike ^a and Joel Howell^b

^aClinical Scholar, National Clinician Scholars Program, Clinical Instructor, Division of Hospital Medicine, Hospitalist, Ann Arbor Va Healthcare System, University of Michigan, Ann Arbor; ^bElizabeth Farrand Collegiate Professor of the History of Medicine, Professor, Departments of Medicine and History, University of Michigan, Ann Arbor

ABSTRACT

The authors conduct a narrative review of the quantitative observation metrics and psychometric scales utilized in the visual arts and medical education literature in order to provide medical educators with a 'toolkit' of quantitative metrics with which to design and evaluate novel visual arts-based pedagogies. These efforts are intended to support the AAMC and National Academy of Sciences, Engineering, and Medicine's aims to formally evaluate and integrate arts and humanities curricula into traditional scientific educational programming. The scales reviewed examine a variety of domains including tolerance for ambiguity, bias, burnout, communication, empathy, grit, and mindfulness/reflection. Observation skill, given the heterogeneity of quantitative metrics, is reviewed separately.

ARTICLE HISTORY

Received 30 August 2021
Revised 27 October 2021
Accepted 20 November 2021

KEYWORDS

Medical education; medical humanities; curriculum development; visual arts; observation skill

Introduction

Since medical humanities' inception in the 1960s, it has grown to encompass a wide range of disciplines, including literature, narrative medicine, music, dance, and others. As of 2018, roughly 85% of medical educational institutions offer some programming in the medical humanities[1]. Starting in the late 1990s, educators began exploring the efficacy of visual arts programming for improving the ability of medical trainees to deliver high-quality healthcare. Most initial studies examined the relationship between visual art programming and skill at visual observation [2,3]. In the years that followed, educators broadened their conception of the potential value of visual arts-based curricula. Some hoped these curricula could help stem rising rates of burnout, others were concerned about changes in medical student empathy, others about tolerance for ambiguity [4–11]. Many educators hoped to demonstrate the value of such training. But how, exactly, does one compute the full impact of engaging with a work of visual art? Educators have attempted to answer this and similar questions by evaluating visual arts programming with a range of tools both qualitative and quantitative [12–16].

Meanwhile, attention to arts and humanities programming by major professional organizations has increased. In 2018, the National Academy of Sciences, Engineering, and Medicine (NASEM) encouraged a renewed engagement with the arts

and humanities in all educational pedagogy[17]. In 2020, the Association of American Medical Colleges (AAMC) called for the formal integration of the arts and humanities into healthcare education and crafted a digital 'Getting Started' guide [18,19]. The AAMC also endorsed a framework for developing medical humanities curricula, the Prism model [20,21]. Both the AAMC and NASEM advocated for novel pedagogical approaches, enhanced research, and a need to measure 'learner outcomes beyond satisfaction with the course or program.' [18]

Some medical humanities educators argue that quantitative measurement of the effects of visual arts-based programming is unnecessary [16,22,23]. But curricular committees may be hesitant to add new medical humanities courses into a crowded curriculum without some form of outcomes-based assessment. While acknowledging the value of qualitative forms of analysis, many leaders in clinical schools may be more comfortable if arguments for visual arts integration are accompanied by some quantitative analysis. As the field matures, scholars may want to compare the impact of curricular programming across different institutions, a task that requires consistent metrics for evaluation. We offer a narrative review of the quantitative observation metrics and psychometric scales utilized in the visual arts and medical education literature in order to provide medical educators with a 'toolkit' of

quantitative metrics with which to design and evaluate novel visual arts-based pedagogies.

Methods – Narrative Review

We conducted an exploratory search of MEDLINE, Web of Science, and Google Scholar for studies that evaluated the role of visual arts in medical education. The search was done in September 2020 with assistance from a health services research librarian. Summative reports, review articles, and bibliographies related to the visual arts and medical education were also reviewed [12–18]. A repeat search of MEDLINE utilizing identical search terms was conducted in April 2021. Inclusion criteria were limited to visual arts-based studies that incorporated quantitative outcomes metrics, were written in English, and conducted in USA and Canadian medical schools or graduate medical education training programs. Studies written before 1980 were excluded. These searches yielded fourteen visual arts and medical education studies [3,24–36]. Strictly qualitative studies are well-summarized elsewhere and were not included in this study [13,15]. This literature review was conducted by one author (JDI) with oversight by the second author (JDH). Below we first address observation and physical diagnosis metrics and then describe the remaining psychometric scales alphabetically (Table 1, Table 2). We do not offer a detailed critique of each metric's design and validation in this narrative review; our intent is to make authors aware of possible assessment techniques as well as their global benefits and limitations. Moreover, as our narrative review of quantitative and psychometric measurements is limited to those used in the visual arts and medical education literature, there are likely other scales that are not explored in this review.

Visual Arts Programming and Medical Education – What Do We Measure?

OBSERVATION AND PHYSICAL DIAGNOSIS

Observation skill is the most frequently assessed quantitative domain in the visual arts and medical education literature (Table 1) [3,25,26,28–33]. Such exploration is logical in that the practice of clinical medicine typically involves practitioners carefully observing patients' bodies. However, as there are no established scales for measuring the accuracy of clinical observation, observation is typically assessed with de novo designed quantitative metrics. These scales often quantify multiple measurements including the number of distinct observations, the number of words used to describe clinical and artistic images, the amount of time spent analyzing an image, or the number of clinically or artistically relevant

Table 1. Observation Scales and Measurements used to Assess Visual Arts Programming.

Quantitative Observation Metrics	Visual Arts-Based Studies
Number of observations (\pm prespecified grading rubric)	(1) Dolev et al. 2001 – Pre-/post-test analysis of clinical images graded on a 9 or 10-point rubric.
Time spent observing an image	(2) Naghshineh et al. 2008 – Pre-/post-test analysis of patient photographs and artistic images using a prespecified rubric. A qualitative analysis of image descriptions was also undertaken.
Number of words used to describe an image	(3) Klugman et al. 2011 – Pre-/post-test analysis of clinical and artistic images by word count and number of observations. (4) Jasani et al. 2013 – Pre-/post-test analysis of patient photographs by number of observations. A qualitative analysis of image descriptions was also undertaken. (5) Klugman et al. 2014 – Pre-/post-test analysis of clinical images by word count and number of observations. (6) Huang et al. 2016 – Pre-/post-test analysis of artistic and dermatologic images by number of observations. (7) Gurwin et al. 2017 – Pre-/post-test analysis of clinical and visual art images. Clinical images were graded on a prespecified rubric. Visual art images were subjectively awarded points for observations, associations, interpretations, flexibility, and empathizing (no prespecified rubric). (8) Goodman et al. 2017 – Pre-/post-test analysis of radiographic images graded on prespecified rubrics. (9) Agarwal et al. 2020 – Pre-/post-test analysis of clinical images by word count and time spent analyzing each image. A qualitative analysis of image descriptions was also undertaken.

observations. However, such scales often involve a fair amount of inherent individual judgment – what, precisely, is a clinically or artistically relevant observation? Heterogeneity between quantitative observation skill metrics has limited inter-study comparison.

AMBIGUITY

The Modified Tolerance for Ambiguity Scale

The Budner Tolerance for Ambiguity (TFA) scale was designed in 1962 and later modified in 1993 for a medical audience by Gail Geller, a professor at Johns Hopkins School of Public Health [6]. The 7-item questionnaire is graded on a Likert Scale and measures an individual's comfort (or discomfort)

Table 2. Psychometric Scales Utilized to Assess Visual Arts Programming (All domains were evaluated using Likert Scales).

Domain	Psychometric Scale	Means Assessed	Visual Arts-based Studies Utilizing Scale(s)
Ambiguity	Modified Tolerance for Ambiguity Scale (TFA)	A 7-item questionnaire that measures an individual's comfort (or discomfort) with ambiguity.	Klugman et al. 2011 Klugman et al. 2014 Gowda et al. 2018 Strohbehn et al. 2020 Gowda et al. 2018
Bias	Best Intentions Questionnaire (BIQ)	A 24-item questionnaire test that assesses how physician biases may affect patient care.	Orr et al. 2019
Burnout	The Maslach Burnout Inventory (MBI)	A 22-item questionnaire that measures three components of burnout in health professional populations: emotional exhaustion, depersonalization, and low sense of personal accomplishment.	
Communication	Communication Skills Attitudes Scale (CSAS)	A 26-item questionnaire that assesses student perceptions towards learning communication skills.	Klugman et al. 2011 Klugman et al. 2014
Empathy & Compassion	The Compassion Scale Inter-reactivity Index (IRI) Jefferson Scale of Empathy Reading the Mind in the Eyes	A 24-item questionnaire that measures compassionate responses through three conflicting domains: kindness-versus-indifferences, common humanity-versus-separation, and mindfulness-versus-disengagement. A 28-item questionnaire that measures affective empathy (empathic concern, personal distress) and cognitive empathy (perspective taking, fantasy). A 20-item questionnaire that measures cognitive and affective components of empathy in health professional populations. Theory of the Mind inspired scale designed to gauge an individual's ability to intuit the emotional state of others through pairing an image of someone's eyes with their internal emotional state.	Zazulak et al. 2015 – IRI scale Zazulak et al. 2017 – IRI scale, Compassion Scale Gurwin et al. 2017 – Reading the Mind in the Eyes Strohbehn et al. 2020 – JSPE
Grit	Short Grit Scale (GRIT-5)	An 8-item questionnaire that measures two components of grit: perseverance of effort and consistency of interest.	Strohbehn et al. 2020
Mindfulness & Reflection	The Five-Facet Mindfulness Scale The Groningen Reflection Ability Scale (GRAS) The Mindful Attention Awareness Scale (MAAS)	A 39-item questionnaire that assesses five domains of mindfulness: description, self-expression, acts of self-awareness, non-judgment of inner experience, and non-reactivity to inner experience. A 23-item questionnaire that measures reflective ability along three domains: self-reflection, empathic reflection, and reflective communication. A 15-item questionnaire that assesses dispositional mindfulness, which is the ability to be aware of and pay attention to the present moment.	Zazulak et al. 2017 – Five Facet Mindfulness Scale Gowda et al. 2018 – GRAS Strohbehn et al. 2020 – MAAS

with ambiguity. Example statements include, 'It really disturbs me when I am unable to follow another person's train of thought' and 'A good task is one in which what is to be done and how it is to be done are always clear.'^[6]

The TFA scale is included in the AAMC Matriculating Student Questionnaire and Graduation Questionnaire administered to all incoming and graduating medical students in the USA^[8]. Intolerance for ambiguity has been found to be associated with increased psychological distress and reduced clinical performance ^[6,8,37,38]. Given the importance of working comfortably in what can often be a fluid and ambiguous clinical world, and considering the ambiguity intrinsic to many works of visual art, the TFA scale has been utilized in several visual art and medical education studies ^[24,27,31,32].

BIAS

Best Intentions Questionnaire

The Best Intentions Questionnaire (BIQ) was designed in 2010 for healthcare trainees by Anne Gill, a Doctor of Nursing in the Department of Pediatrics at Baylor^[39]. The 24-item questionnaire is scored on several Likert Scales and assesses an individual's understanding of their own biases. The first set of statements explores a participant's

perception of how their biases may impact clinical decision making with statements such as 'physicians can have biases about patients about which they are unaware.'^[39] The second half of the questionnaire probes if individuals believe they can learn to become aware of, manage, and eliminate their own biases. Lastly, participants are asked about their ability to recognize their own emotional state in addition to the emotional state of others. One visual arts and medical education study has used the BIQ metric^[27].

BURNOUT

Maslach Burnout Inventory

The Maslach Burnout Inventory (MBI) was developed in 1981 by University of California Berkley psychologist Christina Maslach to measure burnout in a variety of professional groups^[40]. The 22-item questionnaire is scored on a Likert Scale and evaluates three domains of burnout: emotional exhaustion (EE), depersonalization (DP), and sense of personal accomplishment (PA). Responders indicate the frequency they agree with statements such as 'I feel emotionally drained from my work' and 'I don't really care what happens to some patients.'^[40] By this scale, an individual is suffering from symptoms of burnout if they exhibit a high EE score or a high DP score; an individual may also be suffering from

burnout if they exhibit a high EE score plus either a high DP score *or* a low PA score. The prevalence of burnout in the medical population exceeds 67% and it is associated with impaired clinical decision making, malpractice, professionalism lapses, and adverse personal outcomes including substance use and depression [41]. The MBI is recognized by the National Academy of Medicine as one of the most frequently used scale to measure burnout[42]. One 2019 visual arts study utilized the MBI scale to study the effects of an arts intervention on internal medicine residents[34].

COMMUNICATION

Communication Skills and Attitude Test

The Communication Skills and Attitude Test (CSAS) was developed in 2002 by UK psychiatrist Charlotte Rees for measuring medical students' attitudes towards learning communication skills[43]. The 26-item questionnaire is scored on a Likert Scale and is predicated on the belief that to be an effective health-care practitioner, effective communication and an openness to improve one's communication skills are paramount. Example statements include, 'In order to be a good doctor I must have good communication skills' and 'I find it hard to admit to having some problems with my communication skills.'[43] The CSAS has been utilized in two visual art and medical education studies [31,32].

EMPATHY & COMPASSION

The Compassion Scale

The Compassion Scale (CS) was developed in 2011 by psychologist Elizabeth Pommier at the University of Texas at Austin to measure an individual's understanding of and response to the suffering of others [44,45]. The 24-item questionnaire is scored on a Likert scale. It builds upon psychologist and co-creator Kristin Neff's model of self-compassion and measures compassionate responses through three conflicting domains: kindness-versus-indifference, common humanity-versus-separation, and mindfulness-versus-disengagement[44].

Kindness is defined as concern for those who are suffering linked with a desire to console. It is assessed using statements such as 'If I see someone going through a difficult time, I try to be caring towards that person.'[44] Its antithesis, *indifference*, is assessed using statements such as 'I don't concern myself with other people's problems.'[44] *Common humanity* is broadly defined as one's ability to recognize that all humans suffer and simultaneously feel a sense of connection to those who are suffering. It is assessed using statements such as 'Despite my differences with

others, I know that everyone feels pain just like me.'[44] Its antithesis, *separation*, attempts to capture isolation through statements such as 'I can't really connect with other people when they're suffering.'[44] *Mindfulness* is defined as 'balanced awareness that neither avoids nor gets lost in others pain' coupled with a desire to make oneself aware of the other's suffering[44]. Its antithesis, *disengagement*, is an obliviousness to suffering coupled with a lack of desire to offer consolation. Disengagement is measured with statements such as 'I try to avoid people who are experiencing a lot of pain.' [44,45] The scale has been utilized in one visual arts and medical education study[36].

The Interpersonal Reactivity Index

The Interpersonal Reactivity Index (IRI) was developed in 1980 by psychologist Mark Davis at Eckerd College to measure cognitive and affective empathy [46,47]. The 28-item questionnaire is scored on a Likert Scale and measures empathy along four distinct subdomains. *Cognitive empathy* is assessed by examining perspective-taking and fantasy. The perspective-taking statements measure an individual's ability to assume the psychological point of view of others while the fantasy statements examine an individual's ability to imagine themselves as agents in fictional narratives. *Affective empathy* is assessed by examining empathic concern and personal distress. Empathic concern statements evaluate an individual's ability to feel sympathy towards someone who is suffering, while the personal distress statements evaluate an individual's emotional distress that results from witnessing suffering. A shortened version of the IRI has been included in the AAMC Matriculating Student Questionnaire and the Graduating Questionnaire. The IRI has been used in two visual arts-based studies [35,36].

Jefferson Scale of Empathy

The Jefferson Scale of Empathy (JSE) was developed in 2007 by Mohammadreza Hojat, a research professor of psychiatry at Jefferson Medical College, to measure cognitive and affective empathy[48]. The 20-item questionnaire is scored on a Likert Scale and it assesses three components of empathy: perspective taking, compassionate care, and the 'ability to stand in patients' shoes.'[48] The scale is available in 56 different languages and includes modifications for medical students, health professionals, and health professional students. Example statements in the medical student version include, 'patients feel better when their physicians understand their feelings.'[7] This scale has been utilized in one visual arts and medical education study[24].

Reading the Mind in the Eyes

The Reading the Mind in the Eyes test was designed in 1997 by psychologist Simon-Baron Cohen at Cambridge University in the UK to study emotional recognition in patients with Asperger syndrome or high-functioning autism [49–51]. The test is based upon Theory of Mind psychology, which ‘is the ability to recognize the thinking or feelings of others in order to predict their behaviors and act accordingly.’[51] The test asks individuals to match an emotion with images of peoples’ eyes. It is grouped with empathy because the recognition of others’ emotional state is the first step in acting empathetically. Jaclyn Gurwin, an ophthalmologist at the University of Pennsylvania, explored students’ emotional recognition abilities as a secondary outcome with the Reading the Mind in the Eyes test in a visual arts intervention designed to improve first year medical students’ ability to describe retinal and periorbital pathology[28].

GRIT

The Short Grit Scale

The Short Grit Scale (Grit-S) was created in 2009 by Angela Duckworth, a psychologist at the University of Pennsylvania, to measure grit, which is the innate desire to pursue and achieve long-term goals regardless of external positive reinforcement[52]. The 8-item questionnaire is scored on a Likert scale and evaluates grit along two domains: ‘consistency of interest’ and ‘perseverance of effort.’[52] Consistency of interest is assessed using statements such as ‘I have been obsessed with a certain idea or project for a short time but later lost interest.’[52] Perseverance of effort is assessed with statements such as ‘I finish whatever I begin.’[52] This grit scale has been used in one visual arts and medical education study[24].

MINDFULNESS & REFLECTION

The Five-Facet Mindfulness Scale

The Five-Facet Mindfulness Scale was designed in 2006 by University of Kentucky psychologist Ruth Baier to assess mindfulness[53]. The 39-item questionnaire is scored on a Likert Scale and evaluates five distinct domains: observing, describing/self-expression, acting with awareness, non-judgment of inner experience, and non-reactivity to inner experience[53]. *Observing* is assessed using statements such as ‘I pay attention to how my emotions affect my thoughts and behavior.’[53] *Description* and *self-expression* are assessed using statements such as ‘I can usually describe how I feel at the moment in

considerable detail.’[53] *Acts of self-awareness* are assessed using statements such as ‘when I do things, my mind wanders off and I’m easily distracted.’[53] The *non-judgment of inner experience* is assessed using statements such as ‘I criticize myself for having irrational or inappropriate emotions’[53] Lastly, *non-reactivity to inner experience* is assessed with statements such as ‘I perceive my feelings and emotions without having to react to them.’[53] The scale was utilized in one visual arts and medical education study[36].

The Groningen Reflective Ability Scale

The Groningen Reflective Ability Scale (GRAS) was developed in 2009 at the University of Groningen in the Netherlands by Leo Aukes, a researcher at the Center for Research and Innovation of Medical Education, and Joris Slaets, a physician and Professor of Geriatrics[54]. The 23-item questionnaire is scored on a Likert Scale and measures a medical professional’s reflective ability. Predicated on the belief that reflection is required for maintenance of professional competence and personal well-being, the questionnaire evaluates three domains: self-reflection, empathic reflection, and reflective communication.

Self-reflection statements explore an individual’s ability to engage in ‘introspection, exploration, understanding, and appraisal of experiences.’[54] Empathic reflection statements examine the ability to intuit others’ experiences. Reflective communication statements assess one’s openness for feedback and one’s willingness to accept accountability for their actions. The GRAS was utilized in one visual art and medical education study[27].

The Mindful Attention Awareness Scale

The Mindful Attention Awareness Scale (MAAS) was developed in 2003 by University of Rochester psychologists Richard Ryan and Kirk Brown to assesses dispositional mindfulness, which is the ability to be aware of and pay attention to the present moment [55]. Brown and Ryan reinforce that mindfulness is a measure of consciousness distinct from other forms of mental processing, such as cognition or emotion. Consciousness, and by extension mindfulness, requires an *awareness* of the inner and outer environment coupled with a focused *attention* of one’s conscious mind. The 15-item questionnaire is scored on a Likert Scale and it includes statements such as ‘I tend not to notice feelings of physical tension or discomfort until they really grab my attention.’[55] The MAAS was used in a visual arts and medicine study of third year medical students[24].

Discussion: Putting the Toolkit to Use

We surveyed studies of visual art programming and identified several de-novo quantitative scales used to assess observation skill and twelve psychometric scales used to assess a variety of domains: tolerance for ambiguity, bias, burnout, communication, empathy, grit, and mindfulness/reflection (Table 1, Table 2). Some psychometric scales originated in the general psychology literature while others were developed specifically for clinical care and education. Some scales, such as the Jefferson Scale for Empathy and Geller's modified Tolerance for Ambiguity scale have been widely adopted; others, such as the Best Intentions Questionnaire and Reading the Mind in the Eyes test, have gained little traction. The variability of scales used to assess the impact of visual art programming reflects, in part, a lack of consensus regarding how best to measure the utility of visual art interventions for medical trainees. It may also reflect a limited awareness of the universe of available tools with which to evaluate such programming. We hope that this narrative review will make educators planning to evaluate the impacts of visual arts programming aware of the broader universe of analytic tools in order that they can choose the most appropriate one.

Most clinicians and healthcare educators would agree that these metrics attempt to measure domains relevant to medical practice. But many published educational studies do not evidence a clear congruence between their curricular design and the metrics chosen to measure their outcome. It often seems that these scales are indiscriminately incorporated into studies based on ease of administration and interpretation, rather than being carefully selected to match targeted curricular interventions. This practice may also occur, in part, because of a belief that most visual art-based methodologies equally address all the aforementioned domains [12,13,15]. This imprecision has limited our ability to understand the effects of visual arts programming.

Prior to selecting any psychometric tool, study designers should first explore the strengths and weakness of that tool. Consider the psychometric scales used to measure cognitive empathy. In recent years, researchers have attempted to demonstrate quantitatively that empathy changes, and in many cases declines, throughout medical training [7,9–11]. But some argue that the measurement scales used to assess empathy may be flawed given the lack of an agreed-upon operational definition of empathy, the lack of inter-instrument reliability, and, perhaps most important, their reliance on student (or physician) self-assessment [56–61]. Given that 'the preponderance of evidence suggests that physicians have a limited ability to self-assess,' would it not be more

valuable to evaluate empathy through a third-party assessment of clinical encounters with patients (either real or standardized)? [60–62] All of the psychometric scales cited in this review also require individuals to internally judge their agreement (or disagreement) with various statements related to ambiguity, bias, burnout, compassion, communication, and grit. Educators wishing to evaluate novel visual arts programming using psychometric measurement should take this locus of measurement into account before selecting one of these psychometric scales.

How, then, might one use this toolkit? First, educators should attempt to align psychometric scales or quantitative observation metrics with *specific* educational goals. These goals could include one (or more) of the learning objectives outlined in the AAMC-endorsed Prism model: Mastering Skills, Perspective Taking, Personal Insight, and Social Advocacy[20]. For example, an educator who wants to address the 'Personal Insight' domain using a *targeted* visual arts-based intervention may wish to use the psychometric scales that assess tolerance of ambiguity (modified TFA), burnout (MBI), empathy (JSE, IRI), and mindfulness (MAAS). Other psychometric scales not yet utilized in the visual arts and medical education literature may also be useful. For example, those endorsed by the National Academy of Medicine may be especially useful to measure burnout and healthcare professional well-being.⁴²

Second, educators should be aware of the pitfalls of psychometric measurement based on learner self-assessment. Educators may want to pair psychometric measurement with additional data (e.g., qualitative interview, third-party assessment) to evaluate novel curricula more holistically. Finally, educators should consider using these psychometric tools for long-term, longitudinal analysis. While pre-test/post-test measurements may be helpful to gauge the short-term impacts of a visual arts-based curriculum, long-term outcomes related to burnout, professionalism, healthcare worker wellbeing, patient-level outcomes, and other underexplored domains among visual arts programming participants will likely provide much more meaningful data[23]. Carefully matching curricular methodology with psychometric scales, third party assessment, and qualitative data, when accompanied with long term follow-up and reassessment, will build a more cohesive body of medical humanities literature that promotes inter-institution scalability of visual arts programming.

Conclusion

Novel visual arts and humanities pedagogies may need to be formally evaluated to gain recognition

and approval. The NASEM report when coupled with the AAMC monograph and Prism model can support efforts to integrate and disseminate the visual arts into medical education curricula nationwide. This narrative review of measurement scales provides educators with a toolkit of resources to design and evaluate visual arts and medical education initiatives.

Acknowledgments

John David Ike received support for this publication from the VA Office of Academic Affiliations through the VA National Clinician Scholars Program and the University of Michigan Department of Medicine. The contents do not represent the views of the U.S. Department of Veterans Affairs or the U.S. Government.

Disclosure statement

No potential conflict of interest was reported by the author(s).

Funding

This work was supported by the Office of Academic Affiliations, Department of Veterans Affairs

ORCID

John David Ike  <http://orcid.org/0000-0002-6329-753X>

References

- [1] Klugman CM. Medical Humanities Teaching in North American Allopathic and Osteopathic Medical Schools. *J Med Humanit.* 2018;39(4):473–481.
- [2] Bardes CL, Gillers D, Herman AE. Learning to look: developing clinical observational skills at an art museum. *Med Educ.* 2001;35(12):1157–1161.
- [3] Dolev JC, Friedlaender LK, Braverman IM. Use of fine art to enhance visual diagnostic skills. *Jama.* 2001;286(9):1020–1021.
- [4] Dyrbye LN, Thomas MR, Massie FS, et al. Burnout and suicidal ideation among U.S. medical students. *Ann Intern Med.* 2008;149(5):334–341.
- [5] Dyrbye LN, Thomas MR, Shanafelt TD. Systematic review of depression, anxiety, and other indicators of psychological distress among U.S. and Canadian medical students. *Acad Med.* 2006;81(4):354–373.
- [6] Geller G, Tambor ES, Chase GA, et al. Measuring physicians' tolerance for ambiguity and its relationship to their reported practices regarding genetic testing. *Med Care.* 1993;31(11):989–1001.
- [7] Hojat M, Vergare MJ, Maxwell K, et al. The devil is in the third year: a longitudinal study of erosion of empathy in medical school. *Acad Med.* 2009;84(9):1182–1191.
- [8] Geller G, Grbic D, Andolsek KM, et al. Tolerance for Ambiguity Among Medical Students: patterns of Change During Medical School and Their Implications for Professional Development. *Acad Med.* 2021;96(7):1036–1042.
- [9] Hojat M, Mangione S, Nasca TJ, et al. An empirical study of decline in empathy in medical school. *Med Educ.* 2004;38(9):934–941.
- [10] Bellini LM, Shea JA. Mood change and empathy decline persist during three years of internal medicine training. *Acad Med.* 2005;80(2):164–167.
- [11] Bellini LM. Variation of Mood and Empathy During Internship. *JAMA.* 2002;287(23):3143.
- [12] Moniz T, Golafshani M, Gaspar CM, et al. How Are the Arts and Humanities Used in Medical Education? Results of a Scoping Review. *Acad Med.* 2021; Publish Ahead of Print. DOI:10.1097/ACM.0000000000004118.
- [13] Mukunda N, Moghbeli N, Rizzo A, et al. Visual art instruction in medical education: a narrative review. *Med Educ Online.* 2019;24(1):1558657.
- [14] Osman M, Eacott B, Willson S. Arts-based interventions in healthcare education. *Med Humanit.* 2018;44(1):28–33.
- [15] Dalia Y, Milam EC, Rieder EA. Art in Medical Education: a Review. *J Grad Med Educ.* 2020;12(6):686–695.
- [16] Ousager J, Johannessen H. Humanities in undergraduate medical education: a literature review. *Acad Med.* 2010;85(6):988–998.
- [17] National Academies of Sciences E, Medicine. The Integration of the Humanities and Arts with Sciences, Engineering, and Medicine in Higher Education: branches from the Same Tree. Washington DC: The National Academies Press; 2018.
- [18] Howley L, Gauferg E, King B. The Fundamental Role of the Arts and Humanities in Medical Education. *AAMC.* 2020; 1–54 .
- [19] AAMC. Getting Started Guide Part 1: selecting Activities for Integrating the Arts and Humanities into Medical Education. AAMC. <https://www.aamc.org/what-we-do/mission-areas/medical-education/getting-started-activities>. Published 2020. Accessed 2021 May 19
- [20] Moniz T, Golafshani M, Gaspar CM, et al. The Prism Model for Integrating the Arts and Humanities Into Medical Education. *Acad Med.* 2021; Publish Ahead of Print. DOI:10.1097/ACM.0000000000003949.
- [21] Moniz T, Golafshani M, Gaspar CM, et al. The prism model: advancing a theory of practice for arts and humanities in medical education. *Perspect Med Educ.* 2021;10(4):207–214.
- [22] Campo R. A piece of my mind. “The medical humanities,” for lack of a better term. *JAMA.* 2005;294(9):1009–1011.
- [23] Charon R. Commentary: calculating the contributions of humanities to medical practice-motives, methods, and metrics. *Acad Med.* 2010;85(6):935–937.
- [24] Strohbahn GW, Hoffman SJK, Tokaz M, et al. Visual arts in the clinical clerkship: a pilot cluster-randomized, controlled trial. *BMC Med Educ.* 2020;20(1):1.
- [25] Agarwal GG, McNulty M, Santiago KM, et al. Impact of Visual Thinking Strategies (VTS) on the Analysis of Clinical Images: a Pre-Post Study of VTS in First-Year Medical Students. *J Med Humanit.* 2020;41(4):561–572.
- [26] Goodman TR, Kelleher M. Improving Novice Radiology Trainees' Perception Using Fine Art. *J Am Coll Radiol.* 2017;14(10):1337–1340.
- [27] Gowda D, Dubroff R, Willieme A, et al. Art as Sanctuary: a Four-Year Mixed-Methods Evaluation of a Visual Art Course Addressing Uncertainty Through Reflection. *Acad Med.* 2018;93(11S):S8–S13.

- 11SAssociation of American Medical Colleges Learn Serve Lead: Proceedings of the 57th Annual Research in Medical Education Sessions
- [28] Gurwin J, Revere KE, Niepold S, et al. A Randomized Controlled Study of Art Observation Training to Improve Medical Student Ophthalmology Skills. *Ophthalmology*. 2018;125(1):8–14.
- [29] Huang JT, Reynolds SD, DiGiovanni Evans B, et al. Fine arts curriculum improves observational skills of dermatology trainees: a pilot study. *Br J Dermatol*. 2016;175(4):815–817.
- [30] Jasani SK, Saks NS. Utilizing visual art to enhance the clinical observation skills of medical students. *Med Teach*. 2013;35(7):e1327–1331.
- [31] Klugman CM, Beckmann-Mendez D. One thousand words: evaluating an interdisciplinary art education program. *J Nurs Educ*. 2015;54(4):220–223.
- [32] Klugman CM, Peel J, Beckmann-Mendez D. Art Rounds: teaching interprofessional students visual thinking strategies at one school. *Acad Med*. 2011;86(10):1266–1271.
- [33] Naghshineh S, Hafler JP, Miller AR, et al. Formal art observation training improves medical students' visual diagnostic skills. *J Gen Intern Med*. 2008;23(7):991–997.
- [34] Orr AR, Moghbeli N, Swain A, et al. The Fostering Resilience through Art in Medical Education (FRAME) workshop: a partnership with the Philadelphia Museum of Art. *Adv Med Educ Pract*. 2019;10:361–369.
- [35] Zazulak J, Halgren C, Tan M, et al. The impact of an arts-based programme on the affective and cognitive components of empathic development. *Med Humanit*. 2015;41(1):69–74.
- [36] Zazulak J, Sanaee M, Frolic A, et al. The art of medicine: arts-based training in observation and mindfulness for fostering the empathic response in medical residents. *Med Humanit*. 2017;43(3):192–198.
- [37] Caulfield M, Andolsek K, Grbic D, et al. Ambiguity tolerance of students matriculating to U.S. medical schools. *Acad Med*. 2014;89(11):1526–1532.
- [38] Lally J, Cantillon P. Uncertainty and Ambiguity and Their Association with Psychological Distress in Medical Students. *Acad Psychiatry*. 2014;38(3):339–344.
- [39] Gill A, Thompson B, Teal C, et al. Best Intentions: using the Implicit Associations Test to Promote Reflection About Personal Bias. *MedEdPORTAL Publications* 2010; 6:7792
- [40] Maslach C, Jackson SE, Leiter MP. *Maslach burnout inventory manual*. 3rd ed ed. Palo Alto Calif: Consulting Psychologists Press; 1996. 577 College Ave., Palo Alto 94306
- [41] Rotenstein LS, Torre M, Ramos MA, et al. Prevalence of Burnout Among Physicians. *JAMA*. 2018;320(11):1131.
- [42] Dyrbye L, Meyers D, Ripp J, et al. A Pragmatic Approach for Organizations to Measure Health Care Professional Well-Being. *NAM Perspectives*. 2018;8(10). 10.31478/201810b.
- [43] Rees C, Sheard C, Davies S. The development of a scale to measure medical students' attitudes towards communication skills learning: the Communication Skills Attitude Scale (CSAS). *Med Educ*. 2002;36(2):141–147.
- [44] Pommier E, Neff KD, Tóth-Király I. The Development and Validation of the Compassion Scale. *Assessment*. 2020;27(1):21–39.
- [45] Pommier EA. *The Compassion Scale*. US: ProQuest Information & Learning; 2011.
- [46] Davis M. Measuring individual differences in empathy: evidence for a multidimensional approach. *Journal of personality and social psychology*. 1983;44(1):113–126
- [47] Davis MA. *Multidimensional Approach to Individual Differences in Empathy*. JSAS Catalog Sel Doc Psychol. 1980;(10) : 85
- [48] Hojat M. *Empathy in patient care: antecedents, development, measurement, and outcomes*. New York NY: Springer; 2007.
- [49] Baron-Cohen S, Wheelwright S, Hill J, et al. “Reading the Mind in the Eyes” Test Revised Version: a Study with Normal Adults, and Adults with Asperger Syndrome or High-functioning Autism. *J Child Psychol Psychiatry*. 2001;42(2):241–251.
- [50] Baron-Cohen S, Bowen DC, Holt RJ, et al. The “Reading the Mind in the Eyes” Test: complete Absence of Typical Sex Difference in ~400 Men and Women with Autism. *PLOS ONE*. 2015;10(8):e0136521.
- [51] Megías-Robles A, Gutiérrez-Cobo MJ, Cabello R, et al. The ‘Reading the mind in the Eyes’ test and emotional intelligence. *R Soc Open Sci*. 2020;7(9):201305.
- [52] Duckworth AL, Quinn PD. Development and Validation of the Short Grit Scale (Grit-S). *J Pers Assess*. 2009;91(2):166–174.
- [53] Baer RA, Smith GT, Hopkins J, et al. Using Self-Report Assessment Methods to Explore Facets of Mindfulness. *Assessment*. 2006;13(1):27–45.
- [54] Aukes LC, Geertsma J, Cohen-Schotanus J, et al. The development of a scale to measure personal reflection in medical practice and education. *Med Teach*. 2007;29(2–3):177–182.
- [55] Brown KW, Ryan RM. The benefits of being present: mindfulness and its role in psychological well-being. *J Pers Soc Psychol*. 2003;84(4):822–848.
- [56] Colliver JA, Conlee MJ, Verhulst SJ, et al. Reports of the Decline of Empathy During Medical Education Are Greatly Exaggerated: a Reexamination of the Research. *Acad Med*. 2010;85(4):588–593.
- [57] Quince T, Thiemann P, Benson J, et al. Undergraduate medical students' empathy: current perspectives. *Adv Med Educ Pract*. 2016;7:443–455.
- [58] Ward M, Gruppen L, Regehr G. Measuring self-assessment: current state of the art. *Adv Health Sci Educ*. 2002;7(1):63–80.
- [59] Eva K, Regehr G. Self-assessment In The Health Professions: a Reformulation And Research Agenda. *Acad Med*. 2005;80(Supplement):S46–54.
- [60] Davis DA, Mazmanian PE, Fordis M, et al. Accuracy of Physician Self-assessment Compared With Observed Measures of Competence. *JAMA*. 2006;296(9):1094.
- [61] Colliver JA, Verhulst SJ, Barrows HS. EDITORIAL: self-Assessment in Medical Practice: a Further Concern About the Conventional Research Paradigm. *Teach Learn Med*. 2005;17(3):200–201.
- [62] Colliver JA, Willis MS, Robbs RS, et al. Assessment of Empathy in a Standardized-Patient Examination. *Teach Learn Med*. 1998;10(1):8–11.