Impact of Visual Thinking Strategies (VTS) on the Analysis of Clinical Images: A Pre-Post Study of VTS in First-Year Medical Students

Gauri G. Agarwal, Meaghan McNulty, Katerina M. Santiago, Hope Torrents & Alberto J. Caban-Martinez

Journal of Medical Humanities

ISSN 1041-3545

J Med Humanit
DOI 10.1007/s10912-020-09652-4

Springer

Journal of Medical Humanities
https://doi.org/10.1007/s10912-020-09652-4

ORIGINAL RESEARCH

Impact of Visual Thinking Strategies (VTS) on the Analysis of Clinical Images: A Pre-Post Study of VTS in First-Year Medical Students

Gauri G. Agarwal[1] · Meaghan McNulty[1] · Katerina M. Santiago[2] · Hope Torrents[3] · Alberto J. Caban-Martinez[2]

Published online: 21 August 2020

Abstract
To assess the effectiveness of Visual Thinking Strategies (VTS) in medical education curricula, a pretest–posttest experimental study design was used to evaluate the impact of participating in VTS workshops on first-year medical students. A total of forty-one intervention and sixty comparative students completed the study which included the analysis of clinical images followed by a measurement of word count, length of time analyzing images, and quality of written observations of clinical images. VTS training increased the total number of words used to describe clinical images, the time spent analyzing the images, and the number of clinically relevant observations.

Keywords Visual thinking strategies · VTS · Medical humanities · Visual arts in medical education

Introduction

Many medical schools have recognized the role and value of visual arts training and have started to include such training in the medical school curriculum (Perry et al. 2011; Reilly, Ring, and Duke 2005; Klugman, Peel, and Beckmann-Mendez 2011; Naghshineh et al. 2008; Dolev, Friedlaender, and Braverman 2001; Russell 2018; Bardes, Gillers, and Herman 2001; Braverman n.d.; Shapiro, Rucker, and Beck 2006). The use of Visual Thinking Strategies (VTS), a facilitated method of guiding students in analyzing a preselected piece of visual art, has expanded in medical education curricula to develop clinical skills in students and residents.

✉ Gauri G. Agarwal
gagarwal@med.miami.edu

[1] Department of Medicine, University of Miami Miller School of Medicine, 2500 North Military Trail, Suite 260, Boca Raton, FL 33431, USA

[2] Department of Public Health Sciences, Leonard M. Miller School of Medicine, Miami, FL, USA

[3] Lowe Art Museum at the University of Miami, Miami, FL, USA

VTS is grounded in Housen's Theory of Aesthetic Development that posits if learners are exposed to a carefully sequenced series of visual art works, their way of interpreting images change in a predictable manner in which growth in critical and creative thinking accompany growth in aesthetic thought (Housen 2002). VTS involves the use of facilitated small group discussions in which group members are given the opportunity to express their opinions about an artistic piece using three questions: "What is going on in this image?"; "What do you see that makes you say that?"; and "What more can you find?" Facilitators paraphrase the answers of participants and point to the details under discussion (Reilly, Ring, and Duke 2005). Such training can allow students to recognize the power of "slow looking" (turning one's full visual attention to a work of art). In an era of electronic medical records and increasingly complex technology that draws physician's eyes away from patients, such training becomes a potentially important tool in maintaining the gaze on a patient. As Wellbery and McAteer point out, an emphasis on observational skills has "arisen in response to the depersonalization of individuals….and to the efficiencies of practice" (2015). Participation in VTS sessions also requires students to communicate their observations clearly and to listen carefully to the input of others in the group. The use of VTS has also been shown to increase empathy, tolerance of ambiguity, and sense of well-being (Klugman, Peel, and Beckmann-Mendez 2011; Bentwich and Gilbey 2017; Zazulak et al. 2017; Schaff, Isken, and Tager 2011).

There is ample evidence in the literature supporting VTS as a methodology for analysis of visual art (Perry et al. 2011; Reilly, Ring, and Duke 2005; Klugman, Peel, and Beckmann-Mendez 2011; Naghshineh et al. 2008; Dolev, Friedlaender, and Braverman 2001; Russell 2018; Bardes, Gillers, and Herman 2001; Braverman n.d.; Shapiro, Rucker, and Beck 2006). VTS has been utilized at our medical school for several years within a cohort of approximately fifty first-year, dual-degree MD/MPH medical students as part of their first course in medical school, "Introduction to the Medical Profession" (Hailey, Miller, and Yenawine 2015). Students are provided an introductory lecture on the medical humanities including an overview of the VTS methodology, followed by participation in two, three-hour VTS sessions at the university's art museum. Anecdotal feedback and prior course evaluations have indicated that students felt these sessions were effective in enhancing their observational skills, communication skills, and understanding that art, like medicine, does not offer clearly defined answers but rather a multitude of perspectives. Despite the interest in VTS sessions, there was no objective data that such sessions lead to improved observational skills.

There have been a few studies suggesting that VTS can improve observational skills in medical students through the length of time students spent analyzing an image and the number of observations made (Klugman, Peel, and Beckmann-Mendez 2011; Naghshineh et al. 2008; Gurwin et al. 2018). A study by Klugman et al. of thirty-two medical and nursing students across all developmental stages in their education who were trained in VTS indicated that exposure to VTS increased the time spent in observing clinical images, word length, and the number of observations (2011). It was hoped that the increased time and number of observations could translate into observing clinically important findings in patients that would otherwise be missed, but the authors acknowledge that they did not analyze the quality of the observations and did not use a comparative group (Klugman, Peel, and Beckmann-Mendez 2011). Naghshineh et al. conducted a controlled study of fifty-eight pre-clinical students to assess the impact of VTS on the accuracy of observations of clinical images ((2008). In this study, students were also exposed to a significant number of sessions (eight in total) with a didactic component to each session that was distinct from the VTS training which may have also impacted their outcomes. Length of time spent on observations was also not analyzed as

students were provided a pre-determined eight minutes to complete their written descriptions of the clinical images (Naghshineh et al. 2008). Both the Klugman and Naghshineh studies include a self-selected group of student volunteers and used clinical images from physical exams of patients (Klugman, Peel, and Beckmann-Mendez 2011; Naghshineh et al. 2008). Obtaining objective data on the impact of the VTS sessions is important for decision-making regarding the inclusion of VTS in medical education curricula. In the present study, we used a pretest-posttest study design to evaluate the impact of two VTS workshop sessions on the observation skills of first-year medical students. Building on the Klugman and Naghshineh studies, we wanted to assess the impact of VTS alone in a group of non-self selected students with a comparative group and measure word count, amount of time spent in analysis, as well as the quality of observations using clinical images that were not purely physical exam based such as electrocardiogram and radiology images. We hypothesized that medical students who received the VTS training would increase the amount of time they spent analyzing a clinical image, increase the length of their response, and increase the degree of clinically relevant descriptive content.

Methods

Study design

We used a pretest–posttest study design to evaluate the association between participating in a two-session VTS workshop and scores obtained on pretest and a posttest among a sample of first-year medical students in one academic year. Study participants consented and then completed a baseline survey and a pretest. Participants then completed either VTS training (i.e., intervention) or no VTS training (i.e. comparative). After the VTS workshop sessions, the intervention and comparative groups were asked to complete a posttest. The study period was three weeks in length. Institutional Review Board approval was obtained (IRB#20170697).

Study sample, recruitment and consent

First-year medical students from the dual degree MD/MPH program served as the intervention group, and the first-year MD program medical students served as the comparative group within one academic year (Figure 1). Our research team initially provided a short ten-minute outline of the study during the first week of class and provided students with the web link to access the baseline survey instrument. The students were informed that the survey completion was voluntary, and that there would be a $50 incentive at the time of completion of the posttest. The intervention group and the comparative group were recruited separately to minimize potential cross contamination about the intervention within the study sample. All study participants were asked to complete informed consent and a baseline survey via RedCap, an online, secure and encrypted survey administration system.

Visual Training Skills (VTS) workshop intervention sessions and comparative sessions

The VTS workshop consisted of two, three-hour sessions at the University museum over a two-week period. The VTS training sessions began with a half-hour large-group introduction for the intervention group in which two museum educators demonstrated how to co-facilitate a

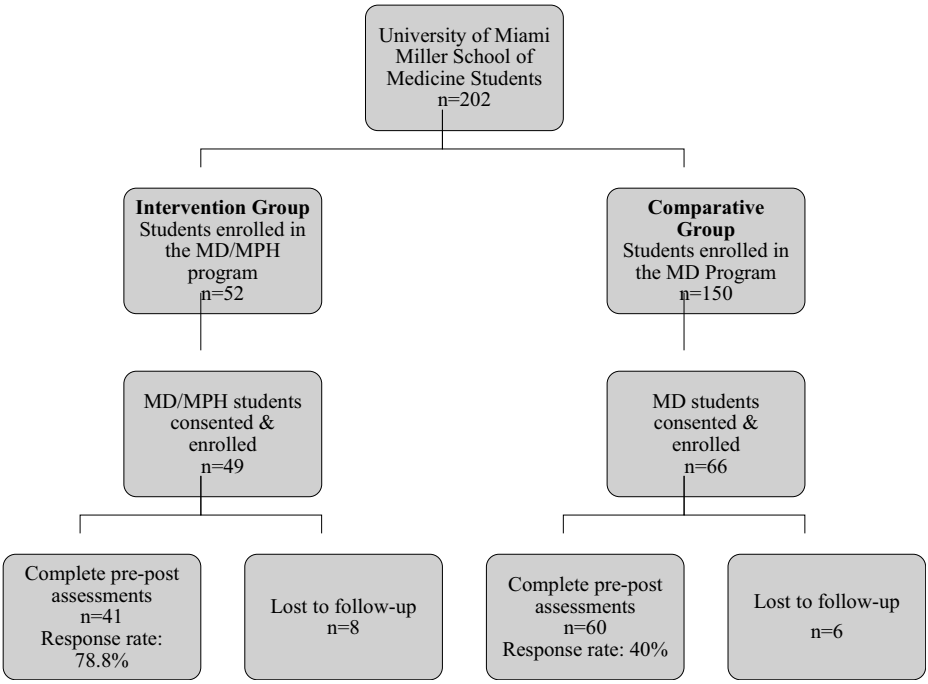

Figure 1. Consort diagram of study sample for the intervention and control group from the University of Miami, Miller School of Medicine first year class, August 2018

VTS discussion. During the introduction, participants were told that they were not only going to participate in the image/object discussions but they would also be asked to co-facilitate. The larger group was then randomly divided into four smaller groups. One museum educator was assigned to lead each group. Participants spent about fifteen minutes discussing a work of art using the VTS methodology. Co-facilitation with the museum educator allowed the participants an opportunity to lead a discussion and be conscious of important elements of VTS: active listening, paraphrasing, linking comments while remaining neutral, and creating an environment that encourages a diversity of ideas and perspectives. A total of six pieces were pre-selected in three different galleries. Each of the four groups looked at a total of two pieces in each gallery. All pieces were pre-selected by a lead museum educator. The content was narrative art with no abstract objects. It was important that there was some ambiguity in each piece which encouraged a diversity of perspectives. Students in the comparative group did not participate in any VTS sessions.

Pretest and posttest assessment measurements

The pretest assessment components were administered prior to the VTS workshop and included: 1) a baseline survey assessing for socio-demographic characteristics and prior clinical or humanities training; and 2) a timed written response portion to clinical images. In the written response section, students were provided with clinical images including normal and abnormal electrocardiograms and chest radiographs as well as a patient with visible physical exam findings indicating an underlying disease process (Jaeger 2015; Kohli 2013; Kim, Lee, and Cho 2013; Deprez et al. 2015; Dow, Yu, and Carmichael 2013). Students were asked to

describe any observations that differed between the normal and abnormal electrocardiograms and chest radiographs and any observations regarding the patient image. The pretest assessment was administered via RedCap where students could type their written responses. A posttest assessment was available to all study participants the day after the last VTS session (for both intervention and comparative groups). Both the intervention and comparative group were sent an automated message via RedCap on the same day to complete a posttest assessment. The posttest assessment did not include a demographics section but did have a timed written response portion structured identically to the pretest assessment.

The free-response answers for the clinical image questions were timed from beginning to completion using the REDCap system. For the pretest questions, the students were presented with a normal and an abnormal electrocardiogram (ECG), a normal and an abnormal chest radiograph, and finally a photo of a patient with Horner's syndrome depicting ptosis and meiosis. For the electrocardiograms and chest radiographs, students were asked to provide written observations of differences noted in the abnormal images. For the final patient image, students were asked to record their observations. For the posttest assessment, the students were similarly timed and were again provided with normal and abnormal electrocardiograms and chest radiographs as well as an image of a patient with Cushing's Syndrome. The images in the pre- and post- test assessments were not identical in that they represented different pathologies but were deemed to be similar in theme by two clinicians on the research team (GA and MM). Differences in response time completion were calculated, as well as differences in word count. Major themes for each set of images (six total) were derived from the response data.

Data analysis

Quantitative analysis: Descriptive analysis was performed on continuous variables and frequencies were generated for categorical variables. Continuous variables were expressed as mean ± the standard deviation of the mean, while categorical variables were expressed as frequency and percent. Statistical analysis of the change in mean thematic observations within each of the groups (i.e., comparative and intervention) were performed using paired t-tests. An Analysis of Variance (ANOVA) was used to analyze the differences between the comparative and intervention group means for each theme at both baseline and post-interventions. All tests were two-sided with an alpha of 0.05 considered statistically significant for all tests. Statistical analysis of survey data was performed with SPSS 24 (IBM Co, Chicago, IL).

Qualitative analysis: A general inductive approach was used to analyze and identify themes within the qualitative data (Thomas 2006). Inductive analysis develops concepts, themes, or a model directly from the raw data through interpretations made by the researcher (Thomas 2006). This approach involved several steps. First, a team of two researchers individually read all of the free responses and summarized key concepts, and they were blinded to which students were in the intervention or comparative group. The team members then preliminarily coded the discussions to generate an initial codebook. The code book contained a list of codes, their corresponding definitions, samples quotes, and decision rules for coding. A second round of coding occurred wherein the pair of raters applied the codes from the code book to free responses. Any coding discrepancies between the pairs were discussed and modifications were made to the codebook. The coders then collapsed codes into broader themes. Consensus was reached among all members to categorize the codes into four broad themes related to the overarching research question. The team worked collaboratively to "score" the free responses by counting the frequency of each theme within individual responses. A total mean score was

created for the comparative and intervention groups by adding the frequency of observations for each image across all images then dividing by the number of students in that group.

Results

Study sample characteristics

A total of one hundred one first-year medical students (forty-one intervention students [40.6%] and sixty comparative students [59.4%]) completed the baseline pretest assessment with no statistical difference in socio-demographic characteristics between the intervention and comparative group with the exception of prior humanities background. Specifically, the intervention student group had a greater proportion of students who identified as having a degree in humanities training than the comparative group (n=13, 31.7% vs n=9, 15.0%; P=0.046, Table 1). Humanities training was assessed using the question, "Please indicate any formal humanities training experiences in which you have participated prior to medical school. The humanities fields usually encompass anthropology, archaeology, classics, history, linguistics and language, literature, performing arts, philosophy, ethics, religion, and visual arts" with the answer choices of "Bachelor's degree in a humanities field; Master's degree in a humanities field; PhD in a humanities field; I am a visual and/or performing artist; Other (please state)." Study completion (both pretest and posttest) was high for both groups, with 90.0% (n=60) of the comparatives and 87.2% (n=41) of the intervention group completing the baseline and posttest assessment.

Qualitative analysis findings

We derived four broad themes from the qualitative data: 1) clinical observations, 2) diagnostic comments, 3) general patient observations, and 4) self-deprecating remarks. We defined "clinical observations" as statements that were descriptive and deemed to be clinically relevant whereas "diagnostic comments" were definitive statements that included a potential diagnosis. "General patient observations" were categorized as remarks made about the image that were neither diagnostic nor clinically relevant but were factually accurate. Many first-year medical students expressed self-doubt regarding their ability to make comparisons and draw conclusions from the patient images, which was detected by word choice and punctuation; these were categorized as "self-deprecating remarks" (Table 2).

Number of clinical observations, diagnostic comments, general patient observations, and "self-deprecating" remarks

At baseline, there were no statistically significant differences between the intervention and comparative group in mean "clinical observations," "diagnostic comments," or "self-deprecating remarks" (p-value= 0.07, 0.45, and 0.21 respectively; Table 3). However, at baseline, the intervention group had a significantly greater mean of "general patient observations" [2.20 (SD=2.64)] compared to the comparative group [1.27 (SD=2.00); P= 0.047].

In the posttest, the intervention group had a significantly greater mean number of "clinical observations" [6.83 (SD=2.90) versus 4.97 (SD=2.40) respectively; P=0.001] and mean number of "general patient observations" [2.63 (SD=2.62) versus 1.62 (2.50) respectively P=0.05] when compared to the comparative group. However, there were no significant

Table 1. Sociodemographic and training characteristics among intervention and comparative group first year medical students participating in the study at the University of Miami (n=101).

Characteristics	Total Sample n (%)†	MD-MPH n (%)†	MD n (%)†	p-value
Total	**101 (100.0)**	**41 (40.6)**	**60 (59.4)**	
Age Group (years)				0.29
Under 21	4 (4.0)	0 (0.0)	4 (6.7)	
21–24	71 (70.3)	28 (68.3)	43 (71.7)	
25–28	24 (23.8)	12 (29.3)	12 (20.0)	
29–34	2 (2.0)	1 (2.4)	1 (1.7)	
Sex				0.83
Male	33 (32.7)	14 (34.1)	19 (31.7)	
Female	68 (67.3)	27 (65.9)	41 (68.3)	
Gender Identity				0.22
Male	32 (31.7)	13 (31.7)	19 (31.7)	
Female	67 (66.3)	26 (63.4)	41 (68.3)	
Gender Queer	2 (2.0)	2 (4.9)	0 (0.0)	
Ethnicity				0.41
Hispanic/Latino	15 (14.9)	7 (17.1)	8 (13.3)	
Non-Hispanic/Latino	85 (84.2)	33 (80.5)	52 (86.7)	
Race				0.28
White	65 (65.0)	30 (75.0)	35 (58.3)	
Black	8 (8.0)	2 (5.0)	6 (10.0)	
Asian	23 (23.0)	6 (15.0)	17 (28.3)	
Other	4 (4.0)	2 (5.0)	2 (3.3)	
Marital Status				0.23
Married	6 (5.9)	1 (2.4)	5 (8.3)	
Divorced, Separated, Widowed	1 (1.0)	1 (2.4)	0 (0.0)	
Single / Living with Partner	94 (93.1)	39 (95.1)	55 (91.7)	
Languages Spoken				0.15
One	31 (31.0)	16 (39.0)	15 (25.4)	
More than one	69 (69.0)	25 (61.0)	44 (74.6)	
Prior Clinical Experiences				
Shadowed Physicians	93 (92.1)	36 (87.8)	57 (95.0)	0.26
Worked as Scribe	21 (20.8)	10 (24.4)	11 (18.3)	0.47
Worked as para-professional	87 (86.1)	38 (92.7)	49 (81.7)	0.12
Medical Interpreter	5 (5.0)	3 (7.3)	2 (3.3)	0.37
Prior Humanities Training				
Degree training (BS, MS, PhD)	22 (21.8)	13 (31.7)	9 (15.0)	**0.046**
Visual/Performing Artist	14 (13.9)	8 (19.5)	6 (10.0)	0.17
Other	18 (17.8)	7 (17.1)	11 (18.3)	0.87
None	69 (68.3)	24 (58.5)	45 (75.0)	0.08
Future Residency Interest				0.11
Primary Care	40 (39.6)	22 (53.7)	18 (30.0)	
Surgical	15 (14.9)	4 (9.8)	11 (18.3)	
Non-Surgical	26 (25.7)	8 (19.5)	18 (30.0)	
Undecided	20 (19.8)	7 (17.1)	13 (21.7)	

†Differences in sub-total population sample due to item non-response or missing.

differences between the intervention and comparative group in the mean number of "diagnostic comments" [0.39 (SD=0.83) versus 0.47 (SD=0.95) respectively; P=0.68] or "self-deprecating remarks" [0.12 (SD=0.33) versus 0.07, (SD=0.25) respectively; P=0.34].

After completing the VTS training sessions, the intervention group made statistically significant more "general patient observations" [2.63 (SD= 2.62)] than they did at baseline [2.20 (SD=2.64); P=0.048]. They also made significantly more "clinical observations" [6.83

Table 2. Themes and example quotations from comparative and intervention first year medical students participating in the study.

Theme	Sample Quotation
Clinical Observations	"The eye on the right side of the image (left eye) has different colors of the iris (green and spotted instead of brown), constricted pupil, swollen eyelids" (Participant 29)
Diagnostic Comments	"One differential diagnosis is Bell's Palsy (paralysis of one side of the face) or a stroke." (Participant 25)
General Patient Observations	"He also has a band-aid on his arm possibly from getting blood drawn or bleeding easily." (Participant 4)
Self-Deprecating Remarks	"My medical knowledge is lacking at this point but the peaks seen in the first row of the first image are diminished in the same row of the second image…" (Participant 56)
	"I don't know how to interpret this." (Participant 90)

(SD= 2.90)] after completing the intervention than at baseline [4.73 (SD=1.76); P < 0.001]. However, there were no statistically significant differences in the mean number of "diagnostic comments" or "self-deprecating remarks" made pre- and post-VTS training for the intervention group. For the comparative group, there were no statistically significant differences in the mean number of any thematic observations made at follow-up when compared to baseline.

Number of words used to describe clinical images

At baseline, the intervention group used statistically significant more words than the comparative group (146.98 words (SD=105.83) versus 91.43 words (SD=64.46), respectively; P=0.001; Table 4) to describe the diagnostic images. This was also demonstrated on the posttest with the intervention group using an average of 172.96 (SD=110.78) words and the comparative

Table 3. Changes in mean frequency of thematic observations for both the intervention and comparative groups at baseline and follow up (n=101)

Theme	Intervention MD-MPH	Control MD	p-value
1. Clinical Observations	**Mean (SD)**	**Mean (SD)**	
Pre-Intervention	4.73 (1.76)	3.85 (2.74)	0.07
Post-Intervention	6.83 (2.90)	4.97 (2.4)	**0.001**
Mean Difference (SD)	**2.01 (3.36)**	1.11 (1.64)	
p-value	**<0.001**	0.07	
2. Diagnostic Comments			
Pre-Intervention	0.46 (0.81)	0.62 (1.09)	0.45
Post-Intervention	0.39 (0.83)	0.47 (0.95)	0.68
Mean Difference (SD)	-0.73 (1.17)	-0.15 (1.19)	
p-value	0.69	0.33	
3. General Patient Observations			
Pre-Intervention	2.20 (2.64)	1.27 (2.00)	**0.047**
Post-Intervention	2.63 (2.62)	1.62 (2.50)	**0.05**
Mean Difference (SD)	**0.44 (1.38)**	0.35 (3.65)	
p-value	**0.048**	0.46	
4. Self-Deprecating Remarks			
Pre-Intervention	0.05 (0.22)	0.13 (0.39)	0.21
Post-Intervention	0.12 (0.33)	0.07 (0.25)	0.34
Mean Difference (SD)	0.73 (0.41)	-0.07 (0.45)	
p-value	0.26	0.25	

group using 98.03 words (SD=73.50; P=<0.001). The intervention group had a statistically significant increase in the number of words used on the posttest when compared to baseline with a mean difference of 25.97 words (SD=80.92, P=0.046). Examples of the students' change in word count and content in their analysis of clinical images are shown in Table 5.

Total time spent writing subjective clinical image observations

At baseline, there was no statistically significant difference in mean completion time between the intervention group [10.66 minutes (SD=6.97)] and comparative group [8.37 minutes (SD=10.99), P=0.24]. However, on the posttest, the intervention group spent a statistically significant greater mean time completing clinical observations than the comparative group [16.63 minutes (SD=23.93) vs. 6.92 minutes (SD=13.34), P=0.01]. Although not statistically significant, the comparative group decreased the amount of time they spent on images [8.37 minutes (SD=10.99) vs. 6.92 minutes (SD=13.34), P=0.52].

Discussion

One of the goals of VTS and of clinical medicine is to train students to expand their visual fields in order to more carefully observe their patients, develop situational awareness, and detail their observations. As with Klugman, we found that VTS increased the mean number of words provided by students to describe clinical images and the time spent analyzing and describing clinical images. The inclusion of a comparative group strengthens these findings. Similar to the Naghshineh study, we found an increase in the number of general observations and the number of clinically relevant observations as compared to a comparative group. Importantly, both of these previous studies have included self-selected student groups, and we wished to see if these outcomes would be seen in students who experienced VTS as part of their required curriculum and with clinical images that were distinct from physical exam findings such as electrocardiogram and radiology images. A prior study with a non-self selected group of one hundred third-year medical students who underwent VTS did not show any difference in the mean number of observations in pre/post testing, but qualitative analysis

Table 4. Changes in number of words used and free response completion time for both the intervention and comparative groups at baseline and follow up (n=101).

	Intervention MD-MPH Mean (SD)	**Control MD Mean (SD)**	**p-value**
Number of Words			
Baseline	146.98 (105.83)	91.43 (64.46)	**0.001**
Post-Intervention	172.96 (110.78)	98.03 (73.50)	**<0.001**
Within Mean Difference (SD)	25.97 (80.92)	6.6 (39.19)	
p-value	**0.046**	0.20	
Free Response Completion Time (minutes)			
Baseline	10.66 (6.97)	8.37 (10.99)	0.24
Post-Intervention	16.63 (23.93)	6.92 (13.34)	**0.01**
Within Mean Difference (SD)	5.97 (22.58)	-1.4 (17.19)	
p-value	0.10	0.52	

Table 5. Growth in word count and content seen in a sample of first year medical students that completed the VTS training sessions.

Pre-Intervention	Post-Intervention
Image B appears to have shorter, wider, and more frequent peaks compared to Image A.	Image B differs from Image A in that the background is lighter red. Additionally, Image B contains some incredibly large dips in the center which are not present in Image A, as well as some double peaks and dips which are not seen in A. Image B does not have any peaks or dips that are as small as some of those seen in Image A.
In image D the lungs appear less clear (more white/gray) than in image C	Image B shows some bright spots down the spinal vertebrae that are looping and do not appear in Image A. Additionally, the dark spot below the heart is larger in Image B than in Image A and the lungs appear cloudier in Image B.
The patient's right eye is lighter in color and speckled. Additionally, the right eyebrow appears higher. The patient has minimal visible eye lashes	The patient has an extremely swollen abdomen, which is not perfectly round but more irregular and bumpy. Patient appears to be male but has swollen pectorial area resembling breast tissue. The center of the stomach has a large scar from the bottom of the ribs to the belly button which is pinkish purple. Veins of the stomach are visible as well as some purple spots near the middle that look like bruises.

did not examine the answers for clinical relevance and the study did not have a comparative group (Jasani and Saks 2013).

The inclusion of this methodology in the required curriculum of intervention and the provision of a supportive grant was critical in achieving a good response rate, as it allowed for a larger sample size and paid participation for both the comparative and intervention groups. At baseline, there appears to be no major difference with regard to demographic features, with the exception of a higher percentage of intervention subjects self-reporting a "humanities" background. We conducted a sensitivity analysis to examine the impact of any humanities training on intervention outcomes (i.e., word count, time spent completing the survey). Overall, we found that prior humanities training whether through formal degree training or informal through work as a performing artist did not significantly impact the total number of words used by trainees in describing the images presented to them as part of the intervention assessment (pre/post images).

This prior exposure may not have necessarily included visual arts training, but in future studies, this should be further detailed at baseline. Despite the increased humanities background, there was no significant difference at baseline in time spent with the images in both groups. Thematic differences were also not noted at baseline, with the exception of a significantly greater number of "general patient observation" comments in the intervention group. It is possible that the previously noted difference in a humanities background led to more general patient observations at baseline. We felt that a background in the humanities may lead to a more analytical mind, already geared for observation and expressiveness.

While more time observing could be considered as inefficient, we feel first-year students, early in their development, should be trained to engage with their patients, taking the time to absorb fully details such as facial expressions, body language, attire, and immediately visible physical exam findings. They can learn how to focus the encounter and work on tailoring their approach to the clinical setting later in their clinical development. The intervention also

appeared to increase the percentage of clinically relevant and general patient observations both within and between groups. Despite a higher baseline number of general observations, there was still a significant increase of observations within the intervention group. There was no statistically significant difference between or within groups of the degree of overt diagnostic comments. The exposure to the intervention was minimal with only two sessions and yet seemed to have an impact.

There are several limitations to this study. The study was carried out at a single institution. In addition, assessment of prior humanities background should be more carefully assessed by adding specific questions on visual arts training. The two groups also had slightly different baseline curricula (comparative students were taking anatomy, and intervention students had taken public health coursework). VTS methodology is also highly dependent on the experience and training of facilitators and in this study we had one dedicated VTS trainer. VTS is currently being used at more than thirty medical schools with varying training of instructors. While there was an increase in clinically relevant observations, it is unknown if this will translate into clinical skills with patients and follow up of students in the clinical setting would be ideal. The design of such a study would be challenging as by the time students are in the clinical environment, it is unlikely that VTS would be the only variable that would impact clinical performance. Lastly, the dosage on the number and frequency of VTS training remains unclear. We selected two three-hour training sessions, but future studies should examine the best dose to deliver VTS training.

The Association of American Medical Colleges (AAMC) is working on a monograph that will include a statement of need and a commitment to integrate the humanities and arts in medical education (AAMC, n.d.). The AAMC is also conducting a scoping review of the use of the humanities and arts in physician training (AAMC, n.d.). This is the first study to our knowledge that assesses the impact of visual arts training on analysis of clinical images beyond the physical exam, utilizing a non-volunteer cohort of first year medical students and a comparative group. There appears to be a clear impact from even short exposure to this type of training and such evidence provides support for greater inclusion of visual arts training in the curriculum at our institution and nationally.

Funding Information This study was funded by Andrew W. Mellon Foundation CREATE Grant (GR009521)Funding.

References

AAMC. n.d. "Association of American Medical Colleges (AAMC)." https://www.aamc.org/what-we-do/mission-areas/medical-education/humanities. By integrating arts and humanities, integration of arts and humanities.

Bardes, C.L., D. Gillers, and A.E. Herman. 2001. "Learning to Look: Developing Clinical Observational Skills at an Art Museum." *Medical Education* 35 (12): 1157–61. http://www.ncbi.nlm.nih.gov/pubmed/11895244.

Bentwich, Miriam Ethel, and Peter Gilbey. 2017. "More than Visual Literacy: Art and the Enhancement of Tolerance for Ambiguity and Empathy." *BMC Medical Education* 17 (1): 200. https://doi.org/10.1186/s12909-017-1028-7.

Braverman, Irwin M. n.d. "To See or Not to See: How Visual Training Can Improve Observational Skills." *Clinics in Dermatology* 29 (3): 343–46. . https://doi.org/10.1016/j.clindermatol.2010.08.001.

Deprez, Fabrice C., Julie Coulier, Denis Rommel, and Antonella Boschi. 2015. "Congenital Horner Syndrome with Heterochromia Iridis Associated with Ipsilateral Internal Carotid Artery Hypoplasia." *Journal of Clinical Neurology (Korea)*. https://doi.org/10.3988/jcn.2015.11.2.192.

Dolev, J.C., L.K. Friedlaender, and I.M. Braverman. 2001. "Use of Fine Art to Enhance Visual Diagnostic Skills." *JAMA* 286 (9): 1020–21. https://doi.org/10.1001/jama.286.9.1020.

Dow, Adrienne, Run Yu, and John Carmichael. 2013. "Too Little or Too Much Corticosteroid? Coexisting Adrenal Insufficiency and Cushing's Syndrome from Chronic, Intermittent Use of Intranasal Betamethasone." *Endocrinology, Diabetes & Metabolism Case Reports*. https://doi.org/10.1530/edm-13-0036.

Gurwin, Jaclyn, Karen E. Revere, Suzannah Niepold, Barbara Bassett, Rebecca Mitchell, Stephanie Davidson, Horace DeLisser, and Gil Binenbaum. 2018. "A Randomized Controlled Study of Art Observation Training to Improve Medical Student Ophthalmology Skills." *Ophthalmology* 125 (1): 8–14. https://doi.org/10.1016/j.ophtha.2017.06.031.

Hailey, Dabney, Alexa Miller, and Philip Yenawine. 2015. "Understanding Visual Literacy: The Visual Thinking Strategies Approach." In *Essentials of Teaching and Integrating Visual and Media Literacy: Visualizing Learning*. https://doi.org/10.1007/978-3-319-05837-5_3.

Housen, Abigail. 2002. "Aesthetic Thought, Critical Thinking and Transfer." *Arts and Learning Research* 18 (1): 2001–2.

Jaeger, Stefan. 2015. "Detecting Disease in Radiographs with Intuitive Confidence." *Scientific World Journal*. https://doi.org/10.1155/2015/946793.

Jasani, Sona K., and Norma S. Saks. 2013. "Utilizing Visual Art to Enhance the Clinical Observation Skills of Medical Students." *Medical Teacher* 35 (7): e1327-31. https://doi.org/10.3109/0142159X.2013.770131.

Kim, Hyun Joo, Hyun Ki Lee, and Belong Cho. 2013. "A Case of Acute Aortic Dissection Presenting with Chest Pain Relieved by Sublingual Nitroglycerin." *Korean Journal of Family Medicine*. https://doi.org/10.4082/kjfm.2013.34.6.429.

Klugman, Craig M, Jennifer Peel, and Diana Beckmann-Mendez. 2011. "Art Rounds: Teaching Interprofessional Students Visual Thinking Strategies at One School." *Academic Medicine : Journal of the Association of American Medical Colleges* 86 (10): 1266–71. https://doi.org/10.1097/ACM.0b013e31822c1427.

Kohli, Rosenman M. 2013. "Indiana University Chest X-Ray Collection." https://openi.nlm.nih.gov/faq.php.

Naghshineh, Sheila, Janet P. Hafler, Alexa R. Miller, Maria A. Blanco, Stuart R. Lipsitz, Rachel P. Dubroff, Shahram Khoshbin, and Joel T. Katz. 2008. "Formal Art Observation Training Improves Medical Students' Visual Diagnostic Skills." *Journal of General Internal Medicine*. https://doi.org/10.1007/s11606-008-0667-0.

Perry, Mark, Nicola Maffulli, Suzy Willson, and Dylan Morrissey. 2011. "The Effectiveness of Arts-Based Interventions in Medical Education: A Literature Review." *Medical Education* 45 (2): 141–48. https://doi.org/10.1111/j.1365-2923.2010.03848.x.

Reilly, Jo Marie, Jeffrey Ring, and Linda Duke. 2005. "Visual Thinking Strategies: A New Role for Art in Medical Education." *Family Medicine* 37 (4): 250–52. http://www.ncbi.nlm.nih.gov/pubmed/15812693.

Russell, Stephen W. 2018. "Improving Observational Skills to Enhance the Clinical Examination." *The Medical Clinics of North America* 102 (3): 495–507. https://doi.org/10.1016/j.mcna.2017.12.011.

Schaff, Pamela B., Suzanne Isken, and Robert M. Tager. 2011. "From Contemporary Art to Core Clinical Skills: Observation, Interpretation, and Meaning-Making in a Complex Environment." *Academic Medicine : Journal of the Association of American Medical Colleges* 86 (10): 1272–76. https://doi.org/10.1097/ACM.0b013e31822c161d.

Shapiro, Johanna, Lloyd Rucker, and Jill Beck. 2006. "Training the Clinical Eye and Mind: Using the Arts to Develop Medical Students' Observational and Pattern Recognition Skills." *Medical Education* 40 (3): 263–68. https://doi.org/10.1111/j.1365-2929.2006.02389.x.

Thomas, David R. 2006. "A General Inductive Approach for Analyzing Qualitative Evaluation Data." *American Journal of Evaluation*. https://doi.org/10.1177/1098214005283748.

Wellbery, Caroline, and Rebecca A, McAteer. 2015. "The Art of Observation: A Pedagogical Framework." *Academic Medicine : Journal of the Association of American Medical Colleges* 90 (12): 1624–30. https://doi.org/10.1097/ACM.0000000000000801.

Zazulak, Joyce, May Sanaee, Andrea Frolic, Nicole Knibb, Eve Tesluk, Edward Hughes, and Lawrence E.M. Grierson. 2017. "The Art of Medicine: Arts-Based Training in Observation and Mindfulness for Fostering the Empathic Response in Medical Residents." *Medical Humanities* 43 (3): 192–98. https://doi.org/10.1136/medhum-2016-011180.