Title
Validating Eye Tracking as an Objective Assessment Tool in Simulation

Journal
CLINICAL SIMULATION IN NURSING, 12(10)

ISSN
1876-1399

Author
Shinnick, MA

Publication Date
2016-10-01

DOI
10.1016/j.ecns.2016.06.001

Peer reviewed

Clinical Simulation in Nursing (2016) 12, 438-446

Clinical Simulation in Nursing

www.elsevier.com/locate/ecsn

Featured Article

Validating Eye Tracking as an Objective Assessment Tool in Simulation

Mary Ann Shinnick, PhD, RN, ACNP*

Assistant Professor In-Residence, University of California at Los Angeles, School of Nursing, Los Angeles, CA 90095, USA

KEYWORDS
simulation;
nursing education;
eye tracking;
nursing assessment;
high stakes testing;
heart failure;
simulation;
point of view video

Abstract

Background: Standardized, objective measures of performance in a simulated experience are lacking. As eye tracking glasses (ETG) provide video, audio, and data capture of a simulation event and quantitative data of participant actions, this technology is well suited for assessment. Therefore, this study sought to begin validating ETG as an adjunct, objective performance assessment tool in simulation.

Method: This was a prospective, validation study with a two-group, convenience sample of novice and expert nurses who participated in a heart failure (HF) simulation scenario to validate ETG using a known-groups approach. A HF scenario designed to elicit seven basic nursing tasks was followed by a knowledge test and demographic questionnaire.

Results: The groups were equivalent in basic HF knowledge as related to the care of a dyspneic patient. Of the seven tasks, all novices completed only one, while in the expert group, all participants completed four of the seven. Significance was found between groups for time to task in five of seven tasks and eye fixation times in key areas.

Conclusions: This pilot study begins the validation process of ETG technology as an objective assessment method as significant differences were elicited between known groups. ETG technology also provides meaningful data and visual images that can be used to inform nursing education in simulation. Additional research is needed to further establish the validity and reliability of ETG technology as an assessment tool in clinical simulation.

Cite this article:
Shinnick, M. A. (2016, October). Validating eye tracking as an objective assessment tool in simulation. *Clinical Simulation in Nursing, 12*(10), 438-446. http://dx.doi.org/10.1016/j.ecns.2016.06.001.

Patient safety relies on health care educators to train and certify safe, competent practitioners. This is especially important as the number of premature deaths related to preventable causes in U.S. hospitals has increased to >400,000 annually (Watling & Lingard, 2012). Efforts to confirm competency in entry-level nursing education include knowledge and skill tests, but there is a lack of standardized assessment. For decades, educators in nursing and medicine have used expert opinion (Watling & Lingard, 2012) and oral examinations (Littlewood, Shilling, Stemland, Wright, & Kirk, 2013), but these types of subjective assessments have problems with interrater reliability and validity (Bensfield, Olech, & Horsley, 2012; Matsell, Wolfish, & Hsu, 1991) and lack standardization between groups and institutions. The use of Objective Structured

* Corresponding author: mshinnic@sonnet.ucla.edu (M. A. Shinnick).

1876-1399/$ - see front matter © 2016 International Nursing Association for Clinical Simulation and Learning. Published by Elsevier Inc. All rights reserved.
http://dx.doi.org/10.1016/j.ecns.2016.06.001

Clinical Examinations (OSCEs) is an improvement over previous assessment methods, but this form of assessment, while objective, is predominately task related (Cazzell & Howe, 2011; Sellu, Davis, & Vincent, 2012) and is not well suited for evaluation of a subject in a simulation which is event based.

Key Points

- Objective measures of performance in simulation are lacking.
- Eye tracking glasses provide a birds eye view of all performance actions.
- Eye tracking glasses may be used with other assessment methods for optimal performance evaluation.

Assessment of performance using simulation may be "low stakes" such as evaluating learning for development reasons or "high stakes" to determine competency. Therefore, since realistic, reproducible situations can now be accurately simulated, the next logical step is to follow the field of aviation in using simulation for testing following formative teaching (Petrusa, 2009). This move has already begun as seen in specialty areas such as anesthesia credentialing (Gallagher & Tan, 2010; Jeffries et al., 2011; Kesten, Brown, & Meeker, 2015; Mudumbai, Gaba, Boulet, Howard, & Davies, 2012).

Despite the popularity of simulation in nursing, there is a large gap in the ability to objectively assess performance during a simulation event. This may be due to the often complex, time sensitive, sequential actions that are difficult to assess, and score (Boulet, Murray, Kras, & Woodhouse, 2008). In addition to a lack of valid and reliable instruments designed for simulation assessment (Mudumbai et al., 2012), other reasons may include the lack of complete visualization of the individual's performance as it is often hampered by the type of equipment used and their placement (microphone placement, camera view, blocked areas or blind spots). Therefore, fine details of student actions could easily be missed, especially if the learner is bent over the "patient" performing a task (i.e., maintenance of sterility in changing a central line dressing or swabbing a port before injection; Henneman & Cunningham, 2005). Problems with audio capture can add frustration in gauging the performance and cause a less than accurate assessment. A possible solution to these issues which may be used as an adjunct to standardized assessment are eye tracking glasses (ETG) as they can overcome many of these difficulties.

Eye tracking technology is based on the features of eye movement and the assumption that eye movements and attention are linked as it traces where attention is being directed (Popa et al., 2015; Rayner, 1998). ETG is a technology already used in other fields such as psychology and marketing research to assess cognition and performance (Jarodzka et al., 2012; Longman, Lavric, & Monsell, 2013; Moacdieh, Prinet, & Sarter, 2013). Eye tracking technology has greatly evolved with imbedded

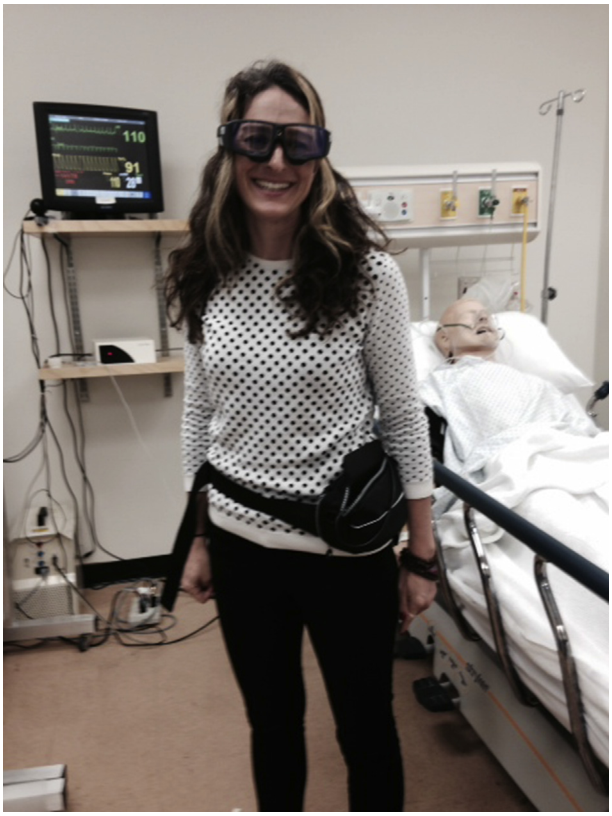

Figure 1 Model wearing eye tracking glasses and fanny pack.

audio and video capture such that blind spots and audio gaps are eliminated. What differentiates ETG from other mobile video sources (i.e., Google Glass), is the extensive, quantitative data capture, and "birds-eye" view of the learner's attention. ETG are now available in lightweight, nontethered glasses (Figure 1). They are also able to produce a video, audio, and eye scanning record of the visual track of the subject's iris with a small "target" placed by the software in the exact location of the subject's gaze. Since the video and visual target follows the subject's gaze and head movements, a "birds-eye" view of all actions and items in the subject's line of sight is available, thus providing an unprecedented observation which even the best simulation center cameras cannot capture. These visual scanning data can be uploaded into computer programs such as Microsoft Excel or IBM SPSS (Bojko, 2013) and provide quantitative data of the visual, cognitive, and attention features of subject performance.

Eye tracking technology has been used in service-related fields such as aviation and medicine to determine competency, improve training, and determine differences between skill levels of pilots, surgeons, and cardiologists (Breen, Bond, & Finlay, 2014; Chetwood et al., 2012; Fisher, Pollatsek, & Pradhan, 2006; Sarter, Mumaw, & Wickens, 2007). Nursing research has reported on students reviewing the ETG simulation video with the goal of performance improvement in subsequent simulations (Henneman & Cunningham, 2005) and in demonstrating differences in visual scanning patterns of nursing students who were able to identify patient errors in identification (Marquard et al., 2011). Studies have shown ETG to be useful in distinct skill assessment in the hospital setting (Ahmidi et al., 2010; Chetwood et al., 2012; Matsumoto et al., 2011; Richstone et al., 2010; Schulz et al., 2011; Voisin, Pinto, Morin-Ducote, Hudson, & Tourassi, 2013) and as an assessment tool in testing novice and expert clinicians (Koh, Park, Wickens, Ong, & Chia, 2011). However, there are no studies testing the validity or reliability of ETG in a simulation setting.

ETG technology can be used as a standardized assessment method in simulation as it produces a variety of data. Data which can be extracted and analyzed after the simulated event include areas of interest (AOIs; identified zones of interest such as the pulse oximeter reading or the ECG tracing on the monitor), the number of fixations on an AOI (where the subject looked, how often and for how long) and the dwell time (time of fixation in an AOI) in those areas, all which can be presented as gaze plots (plot of eye movement) or scan paths. Visual scanning has been used by others as a proxy for determining a subject's focus of attention in complex tasks such as in a study of anesthesiology nurses during surgery to streamline practice (Koh et al., 2011; Seagull, Ward, Mills, Goodrich, & Xiao, 2004). Video capture of vision focal points can be observed in real time with a colored dot or a crosshair on the video. Annotation, or marking, of specific tasks can be done in real time or after event (Bojko, 2013). An additional advantage is the ability to do visual scanning batch analysis on multiple participants at the same time.

As high stakes testing using simulation is becoming more common and desirable in health care, ETG appear well suited as an adjunct for assessment but they have not yet been validated for this purpose. Therefore, the aim of this study was to begin validating ETG for this role by testing the technology in its ability to differentiate clinical performance between known groups (student nurses [novice] and experienced nurses [expert $\geq$5 years of intensive care unit [ICU] experience]) in a heart failure (HF) simulation. A second aim was to explore the utility of the ETG data for skill assessment and response times within a scenario. The hypotheses included (a) the time to complete basic nursing tasks in an HF simulation event (i.e., time to "Applied Oxygen Delivery Device to the Manikin" in a patient complaining of shortness of breath) will be longer in the novice nurses in comparison to the expert nurses and (b) there will be significant fixation differences in AOIs between novice and expert nurses.

Methods

Using a prospective, two-group quasiexperimental design, a convenience sample (n = 35) of novice nurses (senior prelicensure nursing students, n = 16) and expert nurses (ICU nurses of >5 years, n = 19) were recruited from one Baccalaureate School of Nursing (novice nurses/students), a regional tertiary medical center, and a community hospital (expert nurses). Students were introduced to the study by the researcher in one of their classes of which was not being taught by the researcher and nurses were recruited by blanket e-mail to staff in several ICUs. All research activity was done outside of school and work hours. There was no extra credit or grade impact given for participation in this study as the researcher was not one of their educators.

Inclusion criteria for the novice participants were senior undergraduate nursing students who had successfully completed instruction in the care of the decompensated HF patient. For the expert participants, inclusion criteria were status as an ICU or Emergency Department nurse of 5 or more years. Exclusion criteria were participants who either had HF or had family members with HF. The study had university Institutional Review Board approval. No additional IRB approval was needed for hospital nurses as the study was done outside the hospital and on their own time.

Following informed consent and signing of a confidentiality agreement, participants individually were oriented to the ETG, simulation room, manikin, and available supplies (prebrief). After the manufacturer recommended three-point ETG calibration was done so the eye tracker appropriately followed the participant's pupil, each subject participated individually in a 10-minute HF simulation scenario which starts with a patient complaining of dyspnea after eating out at a Mexican restaurant the night before and does not improve until at least oxygen and furosemide (Lasix®) are administered. The objective was not to diagnose the patient but to elicit seven basic nursing tasks: (a) "Applied Pulse Oximeter to Finger," (b) "Observed (looked at) Oximeter Value on Monitor," (c) "Applied Oxygen Delivery Device to the Manikin," (d) "Connected Oxygen Delivery Device to Oxygen (flow meter)," (e) "Auscultated (listened) to Lung Sounds," (f) "Read (checked) Provider's Orders," and (g) "Administered Furosemide" (Lasix). The simulation was immediately followed by a five-question HF Knowledge Test and Demographic questionnaire. Content validation of the simulation and the Knowledge Test was done by cardiac experts, a Cardiologist, and two Cardiac Nurse Practitioners. The Knowledge Test asked both HF- and non-HF—related questions related to nursing care of a dyspneic patient. The objective of this test was to determine if the groups had equivalent knowledge and understanding of HF. The simulation was one of three parallel simulations (different simulations that measure the same construct) that had been pilot tested and used in another large study of >160 nursing students (Shinnick, Woo, Horwich, & Steadman, 2011). The facilitator of all the cases was a trained simulation educator. As the focus of the study was ETG validation between the two groups which was based on the seven tasks, an informal postsimulation debrief was done only if the subject requested it and was done one-on-one after the Knowledge Test. Any postsimulation discussion or debriefing was not part of the study or analysis.

Tetherless ETG from SensoMotoric Instruments (SMI, Germany) were used for this study with video analysis by the SMI Behavioral and Gaze Analysis (SMI BeGaze™) software 3.0. The SMI ETG have been used in numerous fields in multiple studies (SMI, 2016). The ETG are lightweight and support data collection via an Android-based system held in a small pack around the subject's waist ("fanny pack"). Per manufacturer's guidelines, each subject had the ETG calibrated to their pupils individually using a three-point calibration system. The analyzed video did not include time used for calibration or the prebrief, only the simulation time.

Statistical Analysis

A power analyses indicated that a sample size of 34 participants would allow detection of moderate (0.5) effect sizes on an *a priori t* test at an alpha of 0.05 and with a power of 0.80 (Faul, Erdfelder, Lang, & Buchner, 2007). Statistical analysis was done using IBM-SPSS version 20.0 (IBM Corp, 2011) and included parametric statistics (descriptive statistics and independent *t* tests) to compare measurements of time to task between groups and gaze fixation differences (fixation time on task).

Results

An original sample of 46 participants (novices 19; experts 27) was recruited with 11 participants excluded for one of several reasons: no show (6), failure to successfully calibrate the participant's iris (2), incomplete video samples (2), and missing paperwork (1). This allowed for a final sample of 35 participants (novices = 16; experts = 19). While there was a significant difference in age between the groups (novice 25.1 ± 5.6; expert 39.7 ± 10.2; $p < .001$), there was no difference between the groups in the number of simulations they had participated in previously (novice 4.4 ± 1.6; expert 4.3 ± 2.2; $p = .85$) or on a post hoc, five-question, basic knowledge questionnaire of care of a patient who is short of breath (novice 3.7 ± 1.1; expert 4.1 ± 0.9; $p = .21$). This is despite only 50% of novices reportedly caring for a patient complaining of shortness of breath in the clinical setting (Table 1).

Tasks Completed

Of the seven tasks, only one was completed by 100% of all novices, "Checked Provider Orders" while four of the tasks were completed by 100% of the expert participants ("Applied Pulse Oximeter to Finger," "Observed [looked at] Oximeter Value on Monitor," "Applied Oxygen Delivery Device to the Manikin," and "Administered Furosemide"; Table 2).

Table 1 Descriptive Statistics Novices (n = 16) and Experts (n = 19)

	Novices		Experts		
Variable	%	Mean/SD	%	Mean/SD	*p* Value
Age		25.06 ± 5.6		39.68 ± 10.17	<.001
Gender					
Female	81.3		100		
Male	18.7		0		
Number of prior simulations					
1	12.5	4.4 ± 1.6	15.8	4.3 ± 2.2	.85
2	0		21.1		
4	37.4		0		
>4	18.8		5.3		
None	31.3		57.8		
Type of prior hospital employment (novices)					
None	75				
Emergency room technician	12.4				
Admin	6.3				
Information technology	6.3				
Employed in ICU as RN			100		
History of caring for a patient who was short of breath	50		100		
Knowledge Test score (five questions)		3.7 ± 0.08		4.1 ± 0.86	.21

Note. SD = standard deviation.

Table 2 Tasks Expected and Completed Between Groups

Task	Novice or Expert	N	%
Applied pulse oximeter to finger	Novice	15	93.8
	Expert	19	100
Observed oximeter value on monitor	Novice	15	93.8
	Expert	19	100
Applied oxygen delivery device	Novice	15	93.8
	Expert	19	100
Connected oxygen to wall flow meter	Novice	14	93.3
	Expert	18	94.7
Listens to lung sounds	Novice	12	75.0
	Expert	15	78.9
Reads (checks) provider orders	Novice	16	100
	Expert	17	89.5
Administers furosemide	Novice	13	81.3
	Expert	19	100

Time to Task

The time to task for each of the key tasks was computed for each group. The time to task for the novices was slower in all seven tasks but was statistically significant in five (“Applied Pulse Oximeter to Finger” [$p = .05$], “Applied Oxygen Delivery Device to the Manikin” [$p < .01$], “Connected Oxygen Delivery Device to Oxygen (wall flow meter)” [$p < .01$], “Auscultated (listened) to Lung Sounds” [$p = .03$], and “Administered Furosemide” [$p < .01$]). Effect sizes were medium to large on all items (Table 3).

Fixation Differences for Areas of Interest

Fixation data between groups were significant for (a) “Observed (looked at) Oximeter Value on Monitor,” maximum duration ($p = .03$), (b) “Looked at Heart Rate on Monitor,” maximum duration ($p < .01$), and (c) “Read (checked) Provider’s Orders,” number of fixations ($p = .01$; Tables 4-6, respectively). Examples of fixation differences between groups are also illustrated by a scan path plot (Figures 2 and 3).

Participant Feedback

Both expert nurse and student participants were asked about the comfort and wearability of the ETG. Only a few participants wrote comments, and they were not formally analyzed but included, “having the glasses on did not change how I would have taken care of the patient,” “a little tight, however, they felt just like normal sunglasses,” “there could be more padding around the eyes,” and “once the simulation started, I was more concerned about the patient.” The wearability reported by the participants is congruent with other studies actually done in the clinical setting (Koh et al., 2011; Tomizawa, Aoki, Suzuki, Matayoshi, & Yozu, 2012). Importantly, the technology has advanced tremendously since this study was completed with great improvements in comfort and wearability.

Discussion

This pilot study demonstrates the ability of ETG technology to assess differences between known groups in the areas of task completion, time to task completion in most cases, and selected fixations and therefore met Aim 1. The ETG provided data that differentiated groups by substantial fixation differences in AOIs satisfying Aim 2. As in other studies (Koh et al., 2011; Matsumoto et al., 2011; Tomizawa et al., 2012), the scan path images and ETG

Table 3 Time to Task Completion (in Seconds) Between Groups

Task	Novice or Expert	N	Mean/SD ±	p Value	Effect Size
Applied pulse oximeter to finger	Novice	15	83.73 ± 98.39	.048*	0.67 (medium)
	Expert	19	35.16 ± 27.62		
Observed oximeter value on monitor	Novice	16	95.94 ± 95.94	.086	0.58 (medium)
	Expert	19	52.89 ± 52.90		
Applied oxygen delivery device	Novice	15	240.80 ± 240.80	<.001*	1.49 (large)
	Expert	19	84.05 ± 84.05		
Connected oxygen to wall flow meter	Novice	14	235.29 ± 235.29	<.001*	1.44 (large)
	Expert	18	92.22 ± 92.22		
Listens to lung sounds	Novice	14	191.86 ± 191.86	.025*	0.87 (large)
	Expert	15	67.93 ± 67.93		
Reads (checks) provider orders	Novice	16	202.06 ± 202.06	.179	0.50 (medium)
	Expert	18	152.50 ± 152.50		
Administers furosemide	Novice	14	367.50 ± 367.50	.003*	1.1 (large)
	Expert	18	221.28 ± 221.28		

Note. SD = standard deviation.
* Statistically significant at $p < .05$ (some subjects did not complete all of the tasks).

Table 4 Fixation Differences (Areas of Interest; in Seconds) Between Groups: Observed (Looked at) Oximeter Value on Monitor

Looked at Oximeter	Novice or Expert	N	Mean	SD ±	p Value
Total fixation	Novice	16	2,001.9	1,862.2	.24
duration	Expert	14	2,977.6	2,560.2	
Minimal fixations	Novice	16	105.4	119.9	.40
duration	Expert	14	73.3	75.9	
Maximum fixation	Novice	16	335.3	199.1	.03*
duration	Expert	14	602.0	413.5	
Average fixation	Novice	16	175.3	105.8	.10
duration	Expert	14	231.5	68.3	
Total number of	Novice	16	12.9	10.9	.94
fixations	Expert	14	13.2	10.0	

Note. SD = standard deviation.
* Significance < 0.05.

Table 6 Fixation Differences (Areas of Interest; in Seconds) Between Groups: Checks Provider Orders

Checked Orders	Novice or Expert	N	Mean	SD ±	p Value
Total fixation	Novice	10	20,457.5	18,714.4	.03*
duration	Expert	16	8,636.7	6,433.2	
Minimal fixation	Novice	10	61.8	10.17	.16
duration	Expert	16	49.4	25.61	
Maximum fixation	Novice	10	633.1	408.4	.68
duration	Expert	16	577.7	271.0	
Average fixation	Novice	10	162.9	24.74	.07
duration	Expert	16	190.9	41.83	
Total number of	Novice	10	123.0	110.5	.01*
fixations	Expert	16	43.2	27.35	

Note. SD = standard deviation.
* Significance < 0.05.

data revealed a marked difference between how novices and experts are able to process data in a busy environment. While novices spent a significant amount of time looking at data that was not relevant to the "patient's" immediate problem (i.e., focusing on the heart rate of a dyspneic patient), the expert seemed to know what was important and was able to acquire the pertinent information much faster. Other studies have had similar findings in which experts were able to acquire information from an environment as a "chunked set" within shorter periods of time when compared to novices (Kundel, Nodine, Conant, & Weinstein, 2007). The ability to discern and discriminate information that is not pertinent and focus attention to areas of relevance is a skill experts acquire with modeling and experience (Hagemann, Schorer, Canal-Bruland, Lotz, & Strauss, 2010; Nodine, Kundel, Lauver, & Toto, 1996; Savelsbergh, Van der Kamp, Williams, & Ward, 2005) but suggests that nursing educators may need to consider including the development of a student's visual attention in situations requiring clinical reasoning.

Table 5 Fixation Differences (Areas of Interest; in Seconds) Between Groups: Looked at HR on Monitor

Looked at HR	Novice or Expert	N	Mean	SD ±	p Value
Total fixation	Novice	14	720.4	689.6	.55
duration	Expert	13	2,648.5	3,520.1	
Minimal fixations	Novice	14	94.6	58.2	.29
duration	Expert	13	128.7	101.6	
Maximum fixation	Novice	14	218.2	131.5	<.01*
duration	Expert	13	558.8	427.2	
Average fixation	Novice	14	144.3	68.7	.45†
duration	Expert	13	277.7	227.3	
Total number of	Novice	14	4.9	4.7	.18
fixations	Expert	13	11.6	18.0	

Note. HR = heart rate; SD = standard deviation.
* Significance < 0.01.
† Significance < 0.05.

The high demands of health care today require nurses be trained beyond the novice level quicker than ever before. Simulation and the use of eye tracking technology may assist in achieving this goal. Expert nurse performance presented as a scan path or fixation image is very powerful as a model for comparison for novices, so they can gain awareness about and refine their situational focus, noticing skills, and subsequently, clinical performance. In addition, the scan path images can also be used to inform faculty of student weak areas or areas which need more attention or practice (i.e., in a respiratory distress situation, which values on the patient monitor demand the clinician's attention more than others). Fixation data can be useful to convey insight for the novice on areas of relevancy and those that are not pertinent for a given situation. These data provide a unique perspective for the wearer as recall of visual attention is unpredictable. Playback of an expert video or the student's own video may be used to assist novices to a higher level of achievement such as in a study by Henneman et al. (2014). Further study with a detailed debriefing of experts or a simulation using a speak-aloud technique in which participants talk audibly what they are thinking could add more information as to why the expert looked where they looked or to train novices for subsequent simulations. Browning et al. (2015) used the ETG video during debriefing and found it to increase skill level in prelicensure nursing students in subsequent simulations.

An unexpected advantage of using the ETG was the ability to capture quality video and audio with unobstructed views and excellent detail. What differentiates ETG from other mobile video sources (i.e., Google Glass) is the extensive, quantitative data capture, and "birds-eye" view of the learner's attention. There were no blind spots, and all actions were easily visualized, unlike traditional video

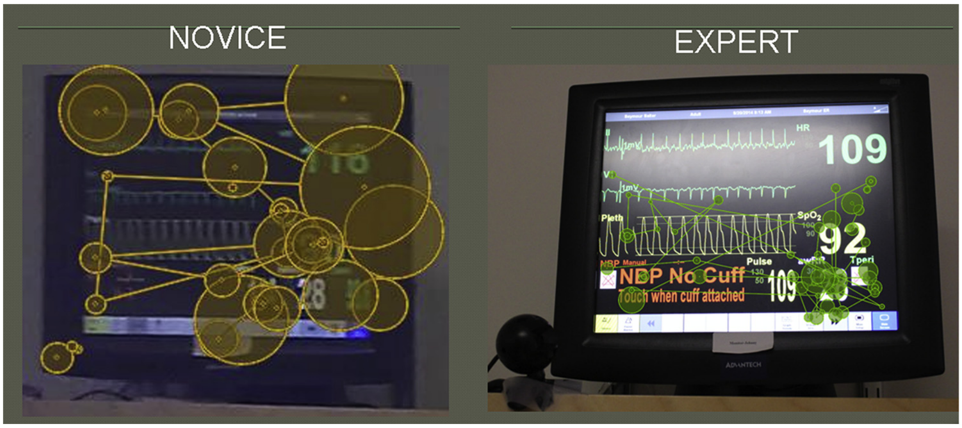

Figure 2 Scan paths for visual ‘‘hits’’ on the monitor. Each circle represents a visual ‘‘hit.’’ The size of the circle represents length of fixation time (i.e., larger circles = longer fixation time).

systems in which fine details of student actions could easily be missed (i.e., swabbing a port before injection; Henneman & Cunningham, 2005).

While not all measures yielded statistical significance between groups, this pilot study begins the validation process of ETG, and those data will result in an objective measure of performance assessment in a simulated experience based on a known-groups validity test. While a larger sample size is desirable, ETG studies generally include smaller sample sizes due to the video analysis required for data retrieval. Sample sizes in other studies range from as low as four (Tomizawa et al., 2012) to twenty four (Matsumoto et al., 2011). While the sample size of this project was large compared with most other ETG studies, further research with larger sample sizes is needed to further establish validity and reliability of ETG as an evaluative tool in a clinical simulation assessment.

Study Limitations

Although efforts were made to minimize study limitations, some were unavoidable. Participants may have had different and unequal clinical experiences in HF before the study. However, there were no statistical differences between the groups on HF knowledge. In addition, while the average simulation experience of the participants was minimal or nonexistent at the time of this study, all were oriented to the simulation manikin and the study environment before the simulation to decrease the effect of differences in simulation experience. A potential concern could be that the subject's awareness of the eye tracking changed their eye movements as they were wearing the glasses, but the participants did not know case details before the study so could not have deliberately altered their focus. Moreover, people cannot control their eye movements to any substantial degree, especially if something is demanding attention (Bojko, 2013). In addition, the eye tracking system used for this study does not require the use of markers in the environment for the ETG to recognize AOIs as some of the other ETG currently being used. The researcher marks AOIs directly on the video following the experiment as a postproduction effort. This is a valuable feature for simulation as it would be difficult to prevent a subject from looking at an area such as the monitor if it had an outline of markers.

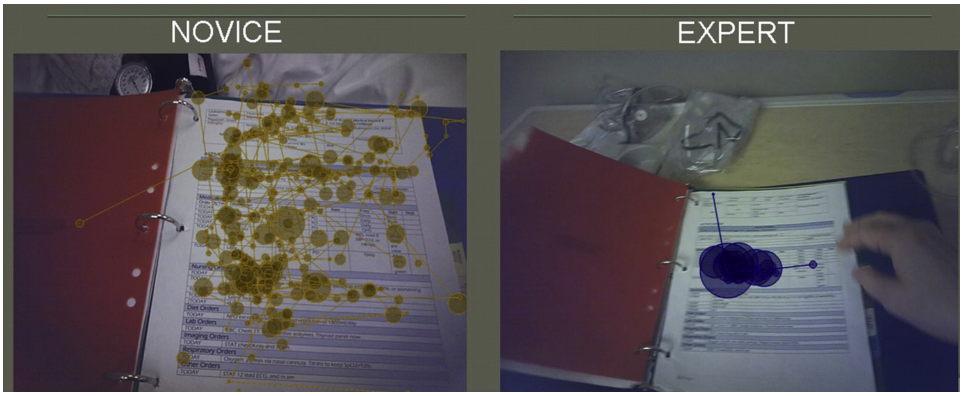

Figure 3 Scan paths for visual ‘‘hits’’ on the provider orders. Each circle represents a visual ‘‘hit.’’ The size of the circle represents length of fixation time (i.e., larger circles = longer fixation time).

A problem with some ETG systems is the loss of gaze focus capture when the subject looks around the glasses (i.e., outside the frame) but the system used in this study (SMI Germany) not only has multiple nose pieces to fit a variety of facial structures for a better fit but also has postevent correction available for this, so minimal amount of data is lost. At the time of this study, some participants who wore glasses had to be eliminated due to the inability of the ETG to calibrate to the subject's pupil. Since this study was completed, improvements have been made to the technology such that participants who wear eye glasses are far more likely able to be calibrated. In addition, the cost of a typical ETG system has been reduced by half. The cost benefit of an ETG purchase and time to train on the system may seem obstructive, but large changes have evolved in this field in the past 2 years.

The use of ETG is promising for the training and assessment of nurses in simulation, most likely as an adjunct to other assessment instruments. This study adds to the current literature on ETG and contributes to the validation process of providing health care professions with an acceptable and objective tool for clinical evaluation.

Acknowledgments

The author would like to acknowledge the other personnel who worked on this project: Dr. Mary A. Woo study design assistance; Christian Cannady, Simulation Technician, University of California at Los Angeles School of Nursing, Nicole Robbins, Simulation Technician University of California at Los Angeles David Geffen School of Medicine, and Megan Guardino, RN, Undergraduate Research Assistant at the University of California at Los Angeles School of Nursing, for their assistance in this study.

References

Ahmidi, N., Hager, G. D., Ishii, L., Fichtinger, G., Gallia, G. L., & Ishii, M. (2010). Surgical task and skill classification from eye tracking and tool motion in minimally invasive surgery. *Medical Image Computing and Computer-Assisted Intervention*, *13*(Pt 3), 295-302.

Bensfield, L. A., Olech, M. J., & Horsley, T. L. (2012). Simulation for high-stakes evaluation in nursing. *Nurse Educator*, *37*(2), 71-74. http://dx.doi.org/10.1097/NNE.0b013e3182461b8c.

Bojko, A. (2013). *Eye Tracking the User Experience: A Practical Guide to Research*. (M. Justak Ed. Vol. 1). Brooklyn, NY: Rosenfeld Media, LLC.

Boulet, J., Murray, D., Kras, J., & Woodhouse, J. (2008). Setting performance standards for mannequin-based acute-care scenarios: An examinee-centered approach. *Simulation in Healthcare*, *3*(2), 72-81.

Breen, C. J., Bond, R., & Finlay, D. (2014). An evaluation of eye tracking technology in the assessment of 12 lead electrocardiography interpretation. *Journal of Electrocardiology*, *47*(6), 922-929. http://dx.doi.org/10.1016/j.jelectrocard.2014.08.008.

Browning, M., Cooper, S., Cant, R., Sparkes, L., Bogossian, F., Williams, B., …, & Black, B. (2015). The use and limits of eye-tracking in high-fidelity clinical scenarios: A pilot study. *International Emergency Nursing*. http://dx.doi.org/10.1016/j.ienj.2015.08.002.

Cazzell, M., & Howe, C. (2011). Using objective structured clinical evaluation for simulation evaluation: Checklist considerations for interrater reliability. *Clinical Simulation in Nursing*, *8*(6), e219-e225.

Chetwood, A. S., Kwok, K. W., Sun, L. W., Mylonas, G. P., Clark, J., Darzi, A., & Yang, G. Z. (2012). Collaborative eye tracking: A potential training tool in laparoscopic surgery. *Surgical Endoscopy*, *26*(7), 2003-2009. http://dx.doi.org/10.1007/s00464-011-2143-x.

Faul, F., Erdfelder, E., Lang, A. G., & Buchner, A. (2007). G*Power 3: A flexible statistical power analysis program for the social, behavioral, and biomedical sciences. *Behavior Research Methods*, *39*(2), 175-191.

Fisher, D. L., Pollatsek, A., & Pradhan, A. (2006). Can novice drivers be trained to scan for information that will reduce their likelihood of a crash? *Injury Prevention*, *12 Suppl 1*, i25-i29.

Gallagher, C. J., & Tan, J. M. (2010). The current status of simulation in the maintenance of certification in anesthesia. *International Anesthesiology Clinics*, *48*(3), 83-99. http://dx.doi.org/10.1097/AIA.0b013e3181eace5e.

Hagemann, N., Schorer, J., Canal-Bruland, R., Lotz, S., & Strauss, B. (2010). Visual perception in fencing: Do the eye movements of fencers represent their information pickup? *Attention, Perception, and Psychophysics*, *72*, 2204-2214.

Henneman, E., & Cunningham, H. (2005). Using clinical simulation to teach patient safety in an acute/critical care nursing course. *Nurse Educator*, *30*(4), 172-177.

Henneman, E., Cunningham, H., Fisher, D., Plotkin, K., Nathanson, B., Roche, J., …, & Henneman, P. (2014). Eye tracking as a debriefing mechanism in the simulated setting improves patient safety practices. *Dimensions of Critical Care Nursing*, *33*(3), 129-135.

IBM Corp. (2011). *IBM SPSS Statistics for Windows, Version 20.0*. Armonk, NY: IBM Corp. Retrieved from www.spss.com.

Jarodzka, H., Balslev, T., Holmqvist, K., Nystrom, M., Scheiter, K., Gerjets, P., & Eika, B. (2012). Conveying clinical reasoning based on visual observation via eye-movement modelling examples. *Instructional Science*, *40*(5), 813-827. http://dx.doi.org/10.1007/s11251-012-9218-5.

Jeffries, P. R., Beach, M., Decker, S. I., Dlugasch, L., Groom, J., Settles, J., & O'Donnell, J. M. (2011). Multi-center development and testing of a simulation-based cardiovascular assessment curriculum for advanced practice nurses. *Nursing Education Perspectives*, *32*(5), 316-322.

Kesten, K. S., Brown, H. F., & Meeker, M. C. (2015). Assessment of APRN student competency using simulation: A pilot study. *Nursing Education Perspectives*, *36*(5), 332-334.

Koh, R. Y., Park, T., Wickens, C. D., Ong, L. T., & Chia, S. N. (2011). Differences in attentional strategies by novice and experienced operating theatre scrub nurses. *Journal of Experimental Psychology Applied*, *17*(3), 233-246. http://dx.doi.org/10.1037/a0025171.

Kundel, H., Nodine, C., Conant, E., & Weinstein, S. (2007). Holistic component of image perception in mammogram interpretation; Gaze-tracking study. *Radiology*, *242*, 396-402.

Littlewood, K. E., Shilling, A. M., Stemland, C. J., Wright, E. B., & Kirk, M. A. (2013). High-fidelity simulation is superior to case-based discussion in teaching the management of shock. *Medical Teacher*, *35*(3), e1003-e1010. http://dx.doi.org/10.3109/0142159x.2012.733043.

Longman, C. S., Lavric, A., & Monsell, S. (2013). More attention to attention? An eye-tracking investigation of selection of perceptual attributes during a task switch. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *39*(4), 1142-1151. http://dx.doi.org/10.1037/a0030409.

Marquard, J. L., Henneman, P. L., He, Z., Jo, J., Fisher, D. L., & Henneman, E. A. (2011). Nurses' behaviors and visual scanning patterns may reduce patient identification errors. *Journal of Experimental Psychology Applied*, *17*(3), 247-256. http://dx.doi.org/10.1037/a0025261.

Matsell, D. G., Wolfish, N. M., & Hsu, E. (1991). Reliability and validity of the objective structured clinical examination in paediatrics. *Medical Education*, *25*(4), 293-299. http://dx.doi.org/10.1111/j.1365-2923.1991.tb00069.x.

Matsumoto, H., Terao, Y., Yugeta, A., Fukuda, H., Emoto, M., Furubayashi, T., …, & Ugawa, Y. (2011). Where do neurologists look when viewing brain CT images? An eye-tracking study involving stroke cases. *PLoS One*, *6*(12), e28928. http://dx.doi.org/10.1371/journal.pone.0028928.

Moacdieh, N. M., Prinet, J. C., & Sarter, N. B. (2013). Effects of modern primary flight display clutter: Evidence from performance and eye tracking data. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, *57*(1), 11-15. http://dx.doi.org/10.1177/1541931213571005.

Mudumbai, S. C., Gaba, D. M., Boulet, J. R., Howard, S. K., & Davies, M. F. (2012). External validation of simulation-based assessments with other performance measures of third-year anesthesiology residents. *Simulation in Healthcare*, *7*(2), 73-80. http://dx.doi.org/10.1097/SIH.0b013e31823d018a.

Nodine, C. F., Kundel, H. L., Lauver, S. C., & Toto, L. C. (1996). Nature of expertise in searching mammograms for breast masses. *Academic Radiology*, *3*(12), 1000-1006. http://dx.doi.org/10.1016/S1076-6332(96)80032-8.

Petrusa, E. R. (2009). Current challenges and future opportunities for simulation in high-stakes assessment. *Simulation in Healthcare*, *4*(1), 3-5. http://dx.doi.org/10.1097/SIH.0b013e3181992077.

Popa, L., Selejan, O., Scott, A., Muresanu, D. F., Balea, M., & Rafila, A. (2015). Reading beyond the glance: Eye tracking in neurosciences. *Neurological Sciences*, *36*(5), 683-688. http://dx.doi.org/10.1007/s10072-015-2076-6.

Rayner, K. (1998). Eye movements in reading and information processing: 20 years of research. *Psychological Bulletin*, *124*(3), 372-422.

Richstone, L., Schwartz, M. J., Seideman, C., Cadeddu, J., Marshall, S., & Kavoussi, L. R. (2010). Eye metrics as an objective assessment of surgical skill. *Annals of Surgery*, *252*(1), 177-182. http://dx.doi.org/10.1097/Sla.0b013e3181e464fb.

Sarter, N. B., Mumaw, R. J., & Wickens, C. D. (2007). Pilots' monitoring strategies and performance on automated flight decks: An empirical study combining behavioral and eye-tracking data. *Human Factors*, *49*(3), 347-357.

Savelsbergh, G. J., Van der Kamp, J., Williams, A. M., & Ward, P. (2005). Anticipation and visual search behaviour in expert soccer goalkeepers. *Ergonomics*, *48*(11-14), 1686-1697. http://dx.doi.org/10.1080/00140130500101346.

Schulz, C. M., Schneider, E., Fritz, L., Vockeroth, J., Hapfelmeier, A., Brandt, T., …, & Schneider, G. (2011). Visual attention of anaesthetists during simulated critical incidents. *British Journal of Anaesthesia*, *106*(6), 807-813. http://dx.doi.org/10.1093/bja/aer087.

Seagull, F., Ward, R., Mills, J., Goodrich, C., & Xiao, Y. (2004). *Measuring awkwardness of workplace layout: Dispersion of attentional and psycomotor resources within the anesthesia workspace*. Paper presented at the Human Factors and Ergonomics Society Annual Meeting Proceedings, New Orleans, Louisiana.

Sellu, D. H., Davis, R. E., & Vincent, C. A. (2012). Assessment of blood administration competencies using objective structured clinical examination. *Transfusion Medicine*, *22*(6), 409-417. http://dx.doi.org/10.1111/j.1365-3148.2012.01192.x.

Shinnick, M., Woo, M., Horwich, T., & Steadman, R. (2011). Debriefing: The most important component in simulation? *Clinical Simulation in Nursing*, *7*(3), e105-e111.

SMI. (2016). *SensoMotoric Instruments*. Retrieved from http://www.smivision.com/en/gaze-and-eye-tracking-systems/home.html.

Tomizawa, Y., Aoki, H., Suzuki, S., Matayoshi, T., & Yozu, R. (2012). Eye-tracking analysis of skilled performance in clinical extracorporeal circulation. *Journal of Artificial Organs*, *15*(2), 146-157. http://dx.doi.org/10.1007/s10047-012-0630-z.

Voisin, S., Pinto, F., Morin-Ducote, G., Hudson, K. B., & Tourassi, G. D. (2013). Predicting diagnostic error in radiology via eye-tracking and image analytics: Preliminary investigation in mammography. *Medical Physics*, *40*(10), 101906. http://dx.doi.org/10.1118/1.4820536.

Watling, C. J., & Lingard, L. (2012). Toward meaningful evaluation of medical trainees: The influence of participants' perceptions of the process. *Advances in Health Sciences Education: Theory and Practice*, *17*(2), 183-194. http://dx.doi.org/10.1007/s10459-010-9223-x.