Designing Serious Games to Train Medical Team Skills

Ashley F. McDermott, James Niehaus
Charles River Analytics, Inc.
Cambridge, MA
amcdermott@cra.com, jniehaus@cra.com

Peter Weyhrauch
Charles River Analytics, Inc.
Cambridge, MA
pweyhrauch@cra.com

ABSTRACT

More access to effective medical team training across the DoD can significantly improve patient outcomes and save lives (Hughes et al., 2016). The DoD and medical communities need implementations of team training protocols that do not require extensive time and personnel commitments. Serious games have lower barriers to enable training because they are highly deployable, are available on-demand, and do not require training staff. This paper reports on the process of implementing team skills training within a single-player serious game. In this paper, we present the process of defining what each of these skills entails, defining the mechanics for how to translate the skills into a game environment, and developing a game that captures these skills. Based on a literature review and interviews with experts, we identified three key team skills that would be our focus for implementation. Once the skills were identified, further work interviewing experts and researching how these skills were measured guided our game design. We encountered several challenges, including identifying appropriate scenarios, choosing appropriate fidelities for the game, and designing a communication platform to support conversation between the user and the virtual teammates. Finally, we validated our design through demonstrations to experts and to medical students, who were the likely audience for the game. This paper provides a case study for how medical team training translated to a game-based training, which will save time, reduce cost, and increase access to the training.

ABOUT THE AUTHORS

Ashley McDermott is a Cognitive Scientist in the Human-Centered Intelligent Systems Division of Charles River Analytics, where she researches and designs learning platforms. These platforms are used to train a range of skills, both cognitive and applied, with applications in mechanical maintenance and medical training. She also has experience in designing game-based training applications and content creation for adaptive learning platforms. Her areas of expertise include learning processes, cognitive skill training, and digital learning platforms. Dr. McDermott earned her Ph.D. in Brain and Cognitive Sciences from the University of Rochester

James Niehaus is a Principal Scientist in the Human-Centered Intelligent Systems Division of Charles River Analytics. Dr. Niehaus' areas of expertise include artificial intelligence and interactive agents. He is currently leading projects to develop psychological resilience applications for the Marines, develop automated adaptive physical training applications, create interactive virtual agents for constructive simulation and training, and develop virtual and adaptive training for first responder trauma assessment. Dr. Niehaus has a Ph.D. in computer science from North Carolina State University.

Peter Weyhrauch is a Principal Scientist and Vice President of the Human-Centered Intelligent Systems Division of Charles River Analytics. Dr. Weyhrauch's primary research interests include modeling and simulation, models of human performance, the cognitive science of learning, artificial intelligence, as well as scenario generation and interactive narrative. He has developed commercial applications of agent-based interactive training and entertainment. Dr. Weyhrauch earned his Ph.D. in Computer Science from Carnegie Mellon University.

Designing Serious Games to Train Medical Team Skills

Ashley F. McDermott, James Niehaus
Charles River Analytics, Inc.
Cambridge, MA
amcdermott@cra.com, jniehaus@cra.com

Peter Weyhrauch
Charles River Analytics, Inc.
Cambridge, MA
pweyhrauch@cra.com

INTRODUCTION

Healthcare organization are emphasizing teamwork, spurred on by the Institute of Medicine's (IOM) report, *To Err is Human: Building a Safer Health System* (Kohn et al., 2000) and reinforced by an updated, alarming estimate of patient harm due to hospital care (James, 2013). The IOM found that systemic, communication-based failures in health care delivery cause more medical errors than individual performance failures, resulting in the IOM's recommendation for enhanced interdisciplinary team training. To address this finding, the Agency for Healthcare Research and Quality (AHRQ) and the Department of Defense, sponsored the development of a medical team training (MTT) resource called TeamSTEPPS (Team Strategies and Tools to Enhance Performance and Patient Safety) (Clancy & Tornberg, 2007) and more recently, TeamSTEPPS 2.0. TeamSTEPPS describes a set of core teamwork principles and provides education on how to develop the necessary skills to achieve successful teamwork in the clinical environment. Several studies have shown MTT can improve clinical metrics including knowledge, communication, error reduction, and patient outcomes (Awad et al., 2005; Merién et al., 2010; Morey et al., 2002).

Unfortunately, even though TeamSTEPPS positively impacts patient safety and outcomes (e.g., Mayer et al., 2011), the extensive time and personnel commitments that TeamSTEPPS training requires are limiting adoption of the framework by medical personnel, both in the DoD and in civilian healthcare. Training for TeamSTEPPS includes classroom exercises and high-fidelity training (e.g., using expensive medical manikins or trained patient role players) to train team skills, as well as serious games. Classroom instruction alone does not prepare trainees completely, and high-fidelity training is costly and difficult because of high teacher-to-student ratios, high cost of training manikins, limited number of scenarios that can be experienced per session, and scheduling demands. Trainees must practice multiple scenarios repeatedly to "proceduralize" their skills. The use of serious games addresses these issues. Serious games have fewer barriers to enable training because they are highly deployable, are available on-demand, and do not require training staff. They can provide a targeted and effective training option that is inexpensive to deploy. However, the current set of serious games do not meet the needs of team training because they are not validated to train medical team skills, they require humans to play each role in the team, and they cannot automatically assess team skills. The DoD and civilian medical communities need serious games for medical team training that: (1) are based on empirical evidence to ensure training effectiveness; (2) are single-player to remove the need to coordinate training across multiple busy medical schedules; and (3) automatically assess team skills to remove the need for extra training staff.

Because game development can be time consuming and expensive, understanding how to design and develop these games so they address the right skills, have the right features, and are customizable is crucial to making these games effective. When serious games are developed on this foundation of understanding, they are an effective training tool that can fit easily into current medical education and training programs. Therefore, we expect that high adoption of effective serious games for training will reduce patient adverse outcomes and deaths caused by medical errors.
This paper will review the process we used to define what medical team skills to include and the key challenges and considerations we faced implementing the training and making the game.

DEFINING TEAM SKILLS

Method

We used TeamSTEPPS training to develop a medical team skills model. We chose TeamSTEPPS because it is a validated measure used to train medical teams and improve communication and leadership skills within the context of a medical team. Based on their materials and the skill model we developed, we held knowledge elicitation interviews with clinical staff in the MedStar Healthcare system, including trauma surgeons and trauma nurses. We used these

interviews to identify critical skills and gain insight into medical team skill training. We cross-validated our findings through a knowledge elicitation session with a subject matter expert (SME) who is an emergency and trauma physician at UMass Memorial Hospital.

We identified three critical skills for medical teams: (1) establishing team and leadership roles; (2) using closed-loop communication; and (3) maintaining team situation monitoring and awareness. Once we identified the skills that would be the focus of the training, we needed to refine the definitions of these skills to implement them in an assessable way within a game environment.

For establishing roles, we interviewed our SME to understand how leadership and roles are typically established in the trauma bay setting. They helped us identify the "huddle" before the team starts working on the patient as the primary time this skill comes into play. We also learned that where a person stands indicates what their role is. For our initial version we included a minimal set of roles consisting of team leader, primary nurse, procedure resident, respiratory technician, airway doctor, and scribe nurse.

For the closed-loop communication, we developed dialogues with feedback from our SME and other trauma doctors affiliated with MedStar Health. Based on the scenario we were using, we identified the key procedures that were necessary to treat the patient in the scenario, including setting up the vitals monitor, necessary steps to intubation, and placing a pelvic binder. The key decision points for those procedures were the subjects for dialogue. We identified what the request from the team leader should be and how the teammates involved in the dialogue should respond if they were effectively closing the loop. From that dialogue we created alternative responses that could be given, including incorrectly repeating what they were told, not responding at all, or asking clarification questions. Once we had the outline for each of these dialogues, we had SMEs review them to make sure all medical jargon was being used correctly.

For situational awareness, we identified multiple frameworks that all categorized team skills differently. None of them directly met our needs, having many questions that were not relevant within the context of our game. To develop a set of situational awareness questions, we combined two of the frameworks. The first was the Endsley model of situational awareness, which divides situational awareness into three levels of comprehension – perception of current situation, comprehension, and ability to project into the future. These capture the level of awareness, but not the particular parts of the situation. The second framework was the Systems Engineering Initiative for Patient Safety (SIEPS), which provided the different categories our situational awareness measures should cover. Based on the combination of these two frameworks, we authored situational awareness questions that would be asked during gameplay.

After defining the team skills, we assessed the usability of our game both with the experts we had interviewed and with a set of medical students who were participating as part of an IRB-approved study. We assessed the usability for two facets in our game design: (1) the effects of different fidelity levels; and (2) the performance of text versus speech inputs. After playing each of the game versions, the experts verbally provided their opinion and the medical students filled out the Post-Study System Usability Questionnaire (PSSUQ).

Results

We identified three critical skills for medical teams: (1) establishing team and leadership roles; (2) using closed-loop communication; and (3) maintaining team situation monitoring and awareness. As we further defined these skills, we decided to create two stages in the game. Based on the interviews with our SME, we developed a team "huddle" stage in which each team member provides their name and a description of what they will be doing during the treatment. The player then must assign official roles to each member of the team. Other than the team lead, all team members were virtual agents. The skill of establishing team roles was assessed through this stage. The second stage was the actual treatment scenario where the game assessed the skills of closed-loop communication and situational awareness. In this scenario, the patient was injured in an auto accident and the trauma team must intubate the patient to stabilize him for surgery. The closed-loop communication dialogues focused on 10 actions that would be necessary to treat the patient. We chose 10 so that we had enough dialogues to capture the different types of responses. Because our users during the project were medical students, we found that we also had to provide them with a checklist, since they were not clear on how to treat the patient. For situational awareness, we developed a total of 15 questions to be asked during gameplay. Each question was tied to a trigger event within the scenario. When triggered, the game would be paused and the question would pop up, obscuring the game until the player responded to the question

In regard to usability, we found that the lower fidelity game was more usable than the medium fidelity medical game (Figure 1). Many issues arose in trying to retrofit a previously existing medium fidelity game that likely had an impact on usability while the low fidelity game was built specifically for this purpose and was designed to be easily changeable. In interviews, users reported preferring speech in theory, but for our implementations there was not a significant difference in usability between text and speech inputs.

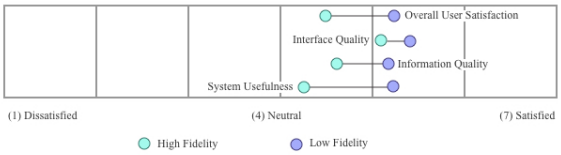

Figure 1. Results from the Post-Study System Usability Questionnaire, showing a preference for low fidelity across all measures.

IMPLEMENTING THE TRAINING

Training Mechanics

How can these skills be implemented in a single-player game?

The largest challenge we encountered during implementation was designing conversations. We created a component that specifically supported conversation, tracking who said what to whom and how they should respond. We kept the possible conversations highly restricted, with a total of ten possible requests to the team from the trauma lead. Each request had many different possible wordings that different doctors use and we tried to capture the most common. The longest conversation included 4 utterances. Being able to track the thread of multiple conversations is critical to this type of game. The game needs to track when a conversation started, whether it ended, and whether the player provided correct responses. Conversations in a medical setting also overlap, so the game needs to track to which conversation the player is responding. Keeping the conversation as restricted as possible helps to keep development scope reasonable, but it can hamper the ability to actually train these skills. Therefore, selecting a minimal set of conversations to train the key skills is important to success.

How does a game provide a believable team with only one human in the loop?

A single-player team skills training game requires well designed virtual teammates. They need to be robust to game events happening out of order or the software having problems understanding commands. Initially we only scripted the planned dialogues, but we soon realized the need for "catch all" statements like "I don't understand what you want me to do" or "Sorry, I'm busy" and to be able to answer the question "What are you doing?"

There is also an argument that it may be better for the teammates to have distinct personalities. When the user has to implement these skills in real life, they will be on teams with other humans who will all have their distinct personalities. We kept personalities to a minimum to help control for conversation complexity but adding personality may improve how much practicing these skills in the game transfers to using them outside of the game.

Assessment

How does the game assess team skills?

For each skill, the game knows what the correct responses and actions are. It compares these with the actions the player takes and creates accuracy scores for each skill. What the game is able to assess depends on what action are available in the game and how they relate to the skills of interest. For example, we kept the available actions to a small number presented in a checklist (Figure 2), so we would not be able to assess medical decision making, but the team members could give different reactions to requests in dialogue, which allowed us to assess closed-loop communication.

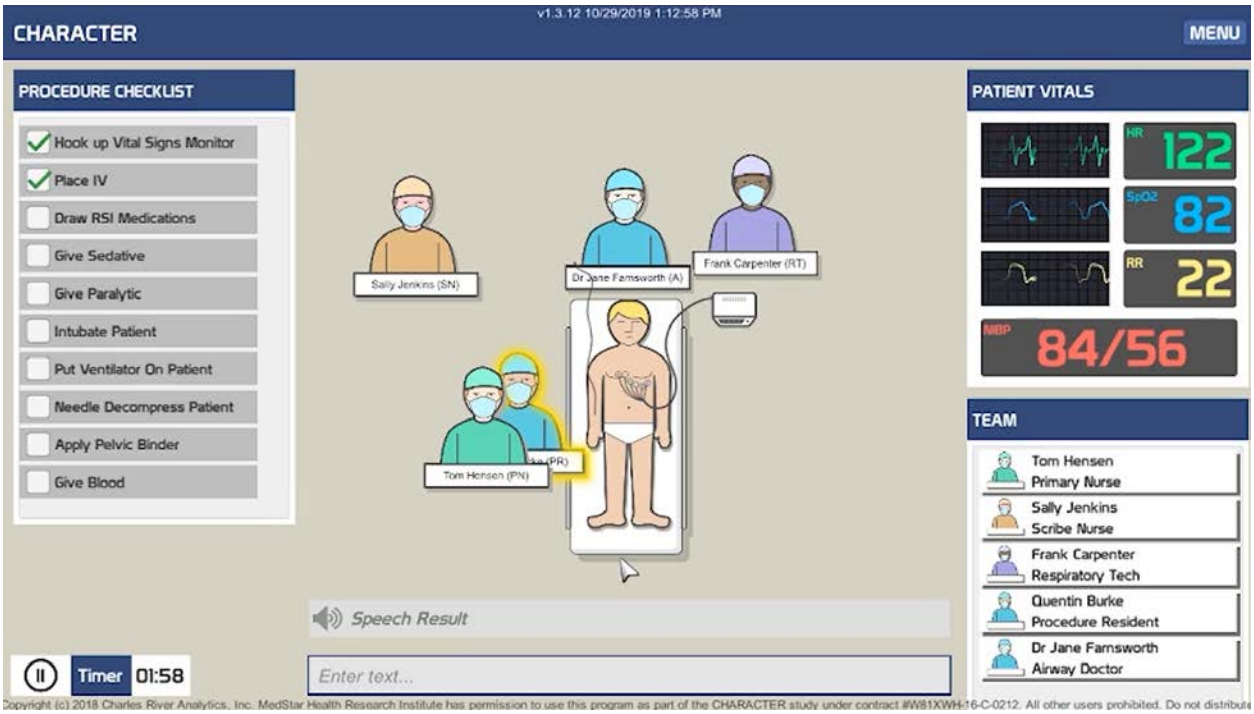

Figure 2. The primary screen during CHARACTER gameplay. A checklist is provided on the left and automatically checks off when a step is completed. Text is entered or the speech results are displayed at the bottom. The team that the user established is listed on the right, as well as vitals information for the patient.

Does the assessment align with traditional assessments?

For potential users to trust that the game is actually training the appropriate team skills, scores from the game should be based on pre-existing, validated measures. Once a validated measure or measures have been selected, then relevant game performance metrics need to be mapped to these measures. We used the Self-Assessment Training Tool for Students (SATTS), which uses a 7-point scale for each item These questions include "Each team member had a clear role" and "When team members received instructions they closed the communication loop", which were assessed from poor (1) to excellent (7). We used standard accuracy measures (the number correct out of total number) for each game performance metric, which produced scores in the form of percentage and then had to map that onto the 7 options. For example, the percentage of correctly closed conversations would be mapped to "When team members received instructions, they closed the communication loop" with 1 = 0-14%; 2 = 15-29%; 3 = 30-44%; 4 = 45-59%; 5 = 60-74%; 6 = 75-89%; 7 = 90-100%.

DESIGNING THE GAME

Game Type

Why use a single-player game?

The primary objective of this effort was to assess critical factors to successfully using a single-player serious game to train team skills. We chose to focus on a single-player game because multi-player games, while not having the same requirement of everyone being physically present at the same place, do still require the difficult task of coordinating schedules so everyone is present at the same time. In many professions, and especially the medical profession, coordinating schedules is a major barrier to training. The experts we interviewed during our field studies all stressed that it was important that this game could be played independently outside of class because medical students could not provide any additional time to coordinate schedules beyond what was already required for medical school. Choosing to design a single-player game to train team skills introduced several key questions that significantly impact the cost and effectiveness of the training system.

Will this have to be its own game or can some of it be reused for other games?

To create an effective medical team training game, the game must have strong communication components that are often lacking in other single-player games. To maximize being able to reuse the work, it is better to develop the game as multiple components that can be pulled out to work in other projects. Building the software as a cohesive whole reduces the possibility for problems in communication between components, but restricts how reusable the software is and how customizable it is. Using the multiple components approach also enables faster customization, such as swapping out text input for speech input. Reuse and modularity in design are key considerations in the engineering of complex software systems such as training games and simulation, and we recommend detailed analysis and careful consideration prior to beginning development because the impact on the project for these two factors is large.

Can a pre-existing medical game be easily modified to add these team skills?
Depending on how the existing game is architected, it may be worth it to create a new game. Early on we compared two possible approaches to the game, either modifying a pre-existing medium fidelity game to train team skills or to create a new low fidelity game. We modified the medium-fidelity trauma medicine game *Trauma Yellow*, adding the communication-based skills on top of it. The game was difficult to modify and could not accommodate high fidelity communication between the player and the virtual teammates. It also required the player to have a high degree of trauma medicine knowledge even though that was not the focus of the training. In the follow-on experiments, we used the low fidelity game that we developed for this purpose. It would be easier to modify a game that already has the communication functionality needed for the training.

Fidelity
Is high fidelity necessary? Is higher fidelity always better?
Fidelity is a broad term that actual covers many different types of fidelity, requiring separate decisions for different features of the game. When deciding what fidelities are most appropriate, it is important to consider who the audience is and what they are supposed to learn. Features that are related to the learning goals should have a reasonable level of fidelity, while features that are not related to the learning goals will generally not require high fidelity. For example, the two games we worked with are presented in Figure 3. The medium fidelity game had higher medical treatment fidelity, which distracted from the team training and led to users having a more difficult time navigating the communication and situational awareness portions of the game. However, higher fidelity can increase engagement, and if it can be achieved for low effort and cost, it is often worth the change. Our low fidelity game had very low fidelity 2D images, but based on recommendations from the medical experts, we added a background noise track and sounds for the vitals monitor, which increased the believability and added an element of stress that would be present when these team skills would actually be used. Our participants responded positively to these features.

Figure 3. Screenshots of *Trauma Yellow*, the medium fidelity game (left) and CHARACTER, the low fidelity game. Both of these games were compared to determine if the low fidelity game could meet the needs of this training.

Does a medical team training game have to be in a medical setting?
Since the team skills are primarily communication-based, it may not be necessary to use a medical setting. TeamSTEPPS training is not conducted in a medical setting. However, the skills we chose to train are straightforward and easy to learn; the primary issue we addressed with our target population has been that professionals do not use them regularly in the trauma bay setting, so we opted to set the game in a trauma bay with a trauma patient who needed immediate treatment. We also received feedback from some medical professionals that the medical setting would be necessary for the player to "buy in" and engage with the training.

Does a medical team training game require input from medical experts?
If the game is taking place in a medical setting, then medical experts should be consulted. It is important to have a high enough fidelity that the game is not providing wildly inaccurate medical care simulation. However, when we interviewed medical experts, they often got caught up in the details of the medical treatment and did not pay any attention to the team skills. We also received conflicting information from different experts and had to adjust the game until it felt reasonable for the players, who were students, not medical experts. Expert input is important to make the medical information generally correct, but it is important to not lose sight of the training goals of the game. We recommend consultation with medical experts while designing medical scenarios and while evaluating intermediate software prototypes for usability and believability.

Interface Design

Can a team skills game work with text input?
Yes, multiple versions of our software used text input (Figure 4) and players showed that they improved their knowledge about team skills. Text is an acceptable way to practice at least some of the communication-based skills targeted with medical team training.

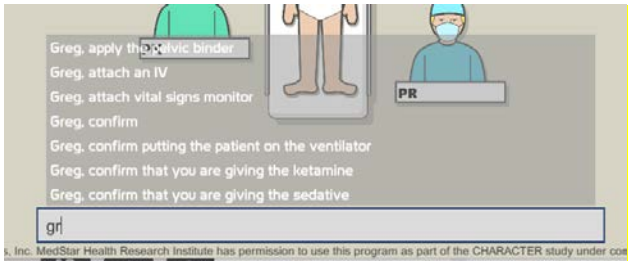

Figure 4. An example of the typing interface. As a user types, they are provided possible options based on what they have already typed. This enabled us to restrict the possible response options while allowing users to find what they want to say quickly and not have to worry about spelling issues.

Is speech input better than text input?
We received feedback from experts and study participants that they greatly preferred speech input to text input. While their preferences are important, our experiment did not show any significant differences between the two interfaces in terms of learning the skills. Therefore, we estimate speech input improves engagement with these systems if it is reasonably well performed with accurate and timely recognition of phrases.

What are the requirements for speech input to work well?
Speech input software continues to improve, but the use case for the game will determine what software can be used and how well it can work. In our experimental setting, internet connectivity was not available, so we had to use the built-in speech-to-text capabilities available in Windows 10. This led to using some unwieldy solutions like pre-processing all of the Agent speech/responses/voice types and storing their audio files and limited the game to only run on Windows 10 machines. Internet connection opens more options, including the ability to use cloud-based natural language processing and artificial intelligence to enable more natural speech communication. Overall, accurate recognition of over 95% of utterances and timely recognition within a couple of seconds seemed to be important to user experience in our studies.

What are the requirements for text input to work well?
There are many ways to communicate the same information and text input is prone to players misspelling words. We addressed this by handling predictive text and text matching for possible user input matching. While we included multiple ways to word a request, ultimately the user still needed to use specific terms to find some of the answers. Text works when the possible inputs can be restricted and response options are easy to search. An alternative would be to allow free typing, but that would require natural language processing and artificial intelligence to understand the intent and adjust for misspellings, and we currently do not see a significant need for this capability beyond what we were able to provide in our experiments.

DISCUSSION AND CONCLUSIONS

In this paper we have presented what we have learned over the course of designing a single-player game to teach medical team skills. We have identified key questions throughout this paper to consider when trying to develop a similar training game as well as how we answered those questions. The paper reviews the process of defining the skills and how to implement the training before designing the game because much of that work should be considered from the beginning of developing training. The key question that has gone unasked in this paper is when is a game appropriate? We found that we could identify skills that were appropriate for practice within a game, but there were

many skills we did not train because they did not fit into the sort of game we were planning to make. Some of these skills may be much better served in a multi-player game or may not lend themselves to being trained in a game at all. The choices around game design should be made as part of the bigger decision of how to appropriately meet users' needs and effectively train the skills of interest. Otherwise you likely will be left with a game that either doesn't train anything or that users won't use. Choosing when it is appropriate to use a game and the best answers to the questions posed here will lead to saving time and money on development, making training time more efficient, and increase the number of people able to participate in the training.

ACKNOWLEDGEMENTS

We would like to thank MedStar's National Center for Human Factors in Healthcare for help recruiting and scheduling the interviews, and Dr. John Broach and trauma specialists who participated in the interviews. This work is supported by the US Army Medical Research and Materiel Command under Contract No. W81XWH-16-C-0212. The views, opinions and/or findings contained in this report are those of the author(s) and should not be construed as an official Department of the Army position, policy or decision unless so designated by other documentation.

REFERENCES

Awad, S. S., Fagan, S. P., Bellows, C., Albo, D., Green-Rashad, B., De La Garza, M., & Berger, D. H. (2005). Bridging the communication gap in the operating room with medical team training. *The American Journal of Surgery*, *190*(5), 770–774. https://doi.org/10.1016/j.amjsurg.2005.07.018

Clancy, C. M., & Tornberg, D. N. (2007). TeamSTEPPS: Assuring Optimal Teamwork in Clinical Settings. *American Journal of Medical Quality*, *22*(3), 214–217. https://doi.org/10.1177/1062860607300616

Hughes, A. M., Gregory, M. E., Joseph, D. L., Sonesh, S. C., Marlow, S. L., Lacerenza, C. N., Benishek, L. E., King, H. B., & Salas, E. (2016). Saving lives: A meta-analysis of team training in healthcare. *The Journal of Applied Psychology*, *101*(9), 1266–1304. https://doi.org/10.1037/apl0000120

James, J. T. (2013). A New, Evidence-based Estimate of Patient Harms Associated with Hospital Care: *Journal of Patient Safety*, *9*(3), 122–128. https://doi.org/10.1097/PTS.0b013e3182948a69

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

Mayer, C. M., Cluff, L., Lin, W.-T., Willis, T. S., Stafford, R. E., Williams, C., Saunders, R., Short, K. A., Lenfestey, N., Kane, H. L., & Amoozegar, J. B. (2011). Evaluating Efforts to Optimize TeamSTEPPS Implementation in Surgical and Pediatric Intensive Care Units. *The Joint Commission Journal on Quality and Patient Safety*, *37*(8), 365-AP3. https://doi.org/10.1016/S1553-7250(11)37047-X

Merién, A. E. R., van de Ven, J., Mol, B. W., Houterman, S., & Oei, S. G. (2010). Multidisciplinary Team Training in a Simulation Setting for Acute Obstetric Emergencies: A Systematic Review. *Obstetrics & Gynecology*, *115*(5), 1021. https://doi.org/10.1097/AOG.0b013e3181d9f4cd

Morey, J. C., Simon, R., Jay, G. D., Wears, R. L., Salisbury, M., Dukes, K. A., & Berns, S. D. (2002). Error Reduction and Performance Improvement in the Emergency Department through Formal Teamwork Training: Evaluation Results of the MedTeams Project. *Health Services Research*, *37*(6), 1553–1581. https://doi.org/10.1111/1475-6773.01104