Developing a tool for observing group critical thinking skills in first-year medical students: a pilot study using physiology-based, high-fidelity patient simulations

Khoa Nguyen, Bertha Ben Khallouq, Amanda Schuster, Christopher Beevers, Nyla Dil, Denise Kay, Jonathan D. Kibble, and David M. Harris
Department of Medical Education, College of Medicine, University of Central Florida, Orlando, Florida

Submitted 25 August 2017; accepted in final form 12 October 2017

Nguyen K, Ben Khallouq B, Schuster A, Beevers C, Dil N, Kay D, Kibble JD, Harris DM. Developing a tool for observing group critical thinking skills in first-year medical students: a pilot study using physiology-based, high-fidelity patient simulations. Adv Physiol Educ 41: 604–611, 2017; doi:10.1152/advan.00126.2017.—Most assessments of physiology in medical school use multiple choice tests that may not provide information about a student’s critical thinking (CT) process. There are limited performance assessments, but high-fidelity patient simulations (HFPS) may be a feasible platform. The purpose of this pilot study was to determine whether a group’s CT process could be observed over a series of HFPS. An instrument [Critical Thinking Skills Rating Instrument CTSRI] was designed with the IDEAS framework. Fifteen groups of students participated in three HFPS that consisted of a basic knowledge quiz and introduction, HFPS session, and debriefing. HFPS were video recorded, and two raters reviewed and scored all HFPS encounters with the CTSRI independently. Interrater analysis suggested good reliability. There was a correlation between basic knowledge scores and three of the six observations on the CTSRI providing support for construct validity. The median CT ratings significantly increased for all observations between the groups’ first and last simulation. However, there were still large percentages of video ratings that indicated students needed substantial prompting during the HFPS. The data from this pilot study suggest that it is feasible to observe CT skills in HFPS using the CTSRI. Based on the findings from this study, we strongly recommend that first-year medical students be competent in basic knowledge of the relevant physiology of the HFPS before participating, to minimize the risk of a poor learning experience.

Introduction

Physiology is the foundation of medicine, but is one of the most difficult topics for medical students to learn. Physiology is conceptual, which represents a challenge for students, since it is difficult to master with rote memorization. Faculty report that students fail to integrate physiological mechanisms across systems and lack the ability to think about dynamic systems in a causal manner (24). These inabilities potentially lead to misconceptions and faulty mental models for medical students. For example, a student deficient in physiological critical thinking (CT) skills may want to treat the tachycardia in a volume-depleted patient instead of the hypotension, because they do not understand the relationship between controlled variables and effectors. These mental models can result in medical errors and compromise patient safety. Therefore, accurate conceptions of, and the ability to think critically about, physiology is essential for the development of clinical reasoning skills.

There are various definitions of CT, but most researchers agree that it requires the application of knowledge, self-reflection, self-correction, and appraising information (2, 30). Walker and Avant (33), nurse educator researchers, determined five attributes that were necessary for CT to occur: knowledge acquisition and application, analysis of information, decision-making, and reflection. A solid base of knowledge is required for knowledge application to more complicated systems (18). Healthcare students need to have the ability to analyze and interpret laboratory values and test results, as well as changes in patient status (29). Students must practice informed decision-making by prioritizing information, weighing evidence-based medicine, and deciding how and when to treat a patient. Reflection allows for students to review decisions and identify possible errors in their thought process that can enhance learning (5). Despite the necessity of CT skills for clinical decision-making, there are limited instruments to measure CT skills. Most approaches for measuring CT have centered on written appraisals, such as the Watson-Glaser Critical Thinking Assessment, which has demonstrated an improvement in CT skills over the 3 yr of medical education (28). Another study using the California Critical Thinking Skills Test showed no improvement in medical science students between semesters (1). CT has also been measured with multiple choice or short-answer tests, but these have met substantial criticism, since test takers are not free to create their own solutions to problems because a single right answer is present (14, 21). Similarly, the majority of assessments during pre-clerkship physician training focus on multiple choice test performance, but, for education to improve, new methods of assessment are needed (7, 20). The Council for Aid to Education proposed that performance assessments have a potential advantage over multiple choice tests because they relate more closely to the real world. Performance assessments that focus on reasoning and information synthesis are less available, but could be important for physiology learning.
In medical education, most of the studies on CT have been in the problem-based learning (PBL) literature. The structure of PBL, where students identify their own learning needs, has been one of the most popular mechanisms in promoting CT skills. Studies have shown that students in PBL are more likely to use meaningful approaches to studying (4, 25). However, many medical schools do not have the number of faculty to support PBL, because each group needs at least one facilitator (34). Another popular way to build CT skills in medical students is through team-based learning (TBL). Supporters of TBL claim that accountability is a key component, since it stimulates the CT process that occurs during the readiness assurance and application phases (23, 31). However, TBL can be difficult for some schools to implement, since it requires a large culture shift and faculty development for successful implementation (26). Other ways to assess and build CT skills are necessary.

High-fidelity patient simulations (HFPS) are a potential modality for performance assessment of physiology CT skills. They contain the four attributes that Walker and Avant (33) determined necessary for CT to occur: knowledge acquisition and application, analysis of information (history, physical examination, laboratory charts, imaging), decision-making (treatment), and reflection (debriefing). However, in our experience, most first-year medical students demonstrate novice behaviors and need substantial guidance during HFPS. The American Dental Education Association (ADEA) outlines the differences between novice and expert behaviors that educators strive to establish in students (17). While the ADEA lists eight common novice and expert behaviors, for brevity, we focused on the three most common behaviors observed in our HFPS. The first common novice behavior is students’ tendency to look for help or “bail outs” as opposed to the expert behavior of taking charge. This includes looking at a faculty facilitator for guidance or assistance. The second novice behavior is slow “trial-and-error” efforts to solve the problem without recognizing that current strategies are not working. This is in opposition to the expert behavior of settling on a “best course of action” quickly, but being willing to change if necessary. The last common novice behavior is focusing on surface features of the problem, as opposed to pinpointing the underlying problem source. Most studies involving HFPS and pre-clerkship medical students have focused on written tests to assess learning, which may not capture the learning process for a student (8, 13, 16). Furthermore, most of the best practices for simulation-based medical education do not focus on novice learners, such as first-year medical students. More studies with first-year medical students and HFPS are necessary to uncover the best approach for assessing CT skills.

There has been limited research on the use of performance assessment to assess basic sciences in first-year medical students. The overall purpose of the present project was to determine the feasibility of observing a group’s physiology CT skills in HFPS. We used IDEAS, a five-step CT problem-solving process as the framework for our performance assessment tool and debriefing (11). The first step, represented by “I,” is Identify problems and set priorities. Step 2, represented by “D,” is Determine relevant information and Deepen understanding. Step 3, represented by “E,” is Enumerate options and anticipate consequences. Step 4, represented by “A,” is Assess situation and make a preliminary decision. Step 5, represented by “S,” is Scrutinize the process and self-correct as needed. The first project aim was to develop a valid and reliable performance assessment instrument for observers to rate group CT skills in a HFPS session. The second aim was to determine whether observable group physiology CT skills improved over a series of HFPS. The overall goal of this project is to begin to develop a “best practices” approach for first-year medical students and HFPS. In addition, having a tool to better identify students who are weaker in CT skills could allow for intervention or improved educational instruction.

METHODS

Study Setting and Participants

This study was reviewed and exempted by the Institutional Review Board of the University of Central Florida. There were 120 first-year medical students in the Class of 2019 who participated in this study. Student groups were formed earlier in the academic year by Student Affairs, with previous medical experience (nurse, physician assistant, military medic) as the first determinant so that there was no more than one in each group. Groups were also formed to take into account sex and ethnicity in the groups. The study occurred during the Structure and Function module (17 wk long from October to February), which is an integrated module consisting of anatomy (microanatomy, embryology, medical imaging) and physiology. Students were broken into 20 groups, with 6 individuals in each. Each group participated in three and observed one HFPS during the module due to time and resource constraints. The HFPS sessions included the following: 1) a basic knowledge quiz and short introduction to the HFPS (10 min); 2) the HFPS session (~9-12 min); and 3) a debriefing session covering physiology objectives (~25 min) and group dynamics (~10 min). Specific information about each portion follows.

Basic knowledge quiz and introduction. Two physiologists and a board-certified clinician created and reviewed each quiz for face and content validity. Every student took a basic knowledge quiz that consisted of 10 true or false questions before each HFPS. The questions covered relevant physiology and microanatomy topics from the preceding week. For this study, only the six to seven physiology questions that pertained directly to the HFPS session were included in the analyses. The score from this quiz is defined as the knowledge score. The introduction, delivered to four groups of students (~24 student) simultaneously, included the setting of the HFPS sessions (urgent care, emergency room, etc.) and information regarding the patient. After the introduction, students proceeded to the simulation center for the HFPS session.

HFPS sessions. SimMan 3G (Laerdal, Norway) manikins were used for the HFPS, and capabilities of the manikins have been previously described elsewhere (8, 16, 19). Due to manikin availability, three groups of students could participate at any given time. One group from the introduction was divided, and two students from this group observed the three participating groups. The clinical scenarios for the four HFPS were as follows: heat exhaustion (week 1), congestive heart failure (week 4), chronic obstructive pulmonary disease (week 6) and diabetic ketoacidosis (week 16). (Table 1). All scenarios were consistent between all groups and did not have interruptions once started. Session instructions directed students to gather and share information from sources such as the manikin (patient), patient monitor and chart, and actors (standardized patients) who were present in the room. Table 1 provides information on each HFPS session, such as the chief problem, relevant underlying physiology, and treatment options. For example, in the heat exhaustion HFPS, students were expected to think about heat-related illness based on vital signs (tachycardia, hyperthermia, hypotension) and a history provided by a standardized patient playing a janitor in the room. The script of the janitor was that the patient was walking outside on a hot day, was lost for hours, and forgot to bring water with him. Students then selected

Advances in Physiology Education • doi:10.1152/advan.00126.2017 • http://advan.physiology.org

Downloaded from http://advan.physiology.org/ by 10.220.33.6 on November 17, 2017
a treatment option for the hyperthermia (cold packs) and hypotension [intravenous (IV) fluids]. To choose the correct IV fluid, students needed to understand the osmotic differences in the three fluid options. All HFPS were video recorded by a CAE Healthcare software for video raters. Following the simulation center, students proceeded to a debriefing room.

Each group had a facilitator, who served in the role of nurse, to control for technical aspects, such as inserting the IV, raising the bed, or applying an oxygen mask. The facilitators consisted of two physiologists, two clinicians, and two fourth-year medical students. Facilitators met as a group before each simulation to test programming of the manikin and standardize the approach to facilitating each group during the HFPS.

Debriefing sessions. Following the HFPS session, the debriefing session consisted of two parts. All of the groups convened for the first part, when faculty debriefed the physiology objectives for each case (~25 min). A physiologist (D.H. or J.K.) and a clinician cofacilitated each debriefing session. The items listed on the CT Skills Rating Instrument (CTSRI) serve as prompting questions for the debriefing session (Fig. 1). Next, the groups participated individually for the second part of the debriefing session with their faculty facilitator (~10 min). The discussion for the second part of the debriefing session related to group dynamics and teamwork.

**CTSRI**

The CTSRI was developed and validated for content by three investigators (D.H., J.K., and D.K.) (Fig. 1). The CTSRI was designed based on the five-step CT problem-solving process called IDEAS, described in the introduction. Each step or part of the process had one observation, except for the second step of “Determine,” which had two. For this study, we define each step as an observation. The CTSRI allows for the rating of six observable behaviors of CT on a five-point Likert scale (0–4), with scale response anchors that are observation specific. For example, for Identify, a rating of 0 was “Not able to identify chief problem(s) of the patient” and denoted low CT skills. On the other hand, a rating of 4 was “ Able to identify chief problem(s) and was able to prioritize key issues without prompting,” which denoted high CT skills. Internal consistency of the CTSRI was good (see RESULTS).

Table 1. Descriptions of chief problem(s) and relevant underlying physiology and treatment options for the four high-fidelity patient simulations

<table>
<thead>
<tr>
<th>Simulation</th>
<th>Chief Problem(s)</th>
<th>Relevant Underlying Physiology</th>
<th>Treatment Options</th>
</tr>
</thead>
<tbody>
<tr>
<td>Heat exhaustion</td>
<td>Hyperthermia, hypotension</td>
<td>1. Osmotic differences between IV fluids.</td>
<td>Cold packs, heat blanket, 0.9% normal saline, 0.45% normal saline, 5% dextrose</td>
</tr>
<tr>
<td>Congestive heart failure</td>
<td>Hypotension, decreased cardiac contractility</td>
<td>1. Controlled variables (blood pressure) vs. effectors (heart rate).</td>
<td>0.9% Normal saline, nitropresside (vasodilator), dobutamine (β-adrenergic agonist), propanolol (β-adrenergic antagonist), epinephrine (α1-adrenergic agonist), vasopressin (vasoconstrictor, anti-diuretic hormone)</td>
</tr>
<tr>
<td>Chronic obstructive pulmonary disease (patient has allergy to penicillin)</td>
<td>Hypoxia</td>
<td>1. Oxyhemoglobin dissociation curve.</td>
<td>Albuterol (nebulized), ipratropium (anti-cholinergic, nebulized), dobutamine, propanolol, epinephrine, penicillin</td>
</tr>
<tr>
<td>Diabetic ketoacidosis (acute phase)</td>
<td>Hypotension, ketoacidosis</td>
<td>1. Actions of insulin.</td>
<td>0.9% Normal saline, 0.45% normal saline, 5% dextrose, insulin, potassium, sodium bicarbonate</td>
</tr>
</tbody>
</table>

Correct options are in bold.
Instructions for Faculty Member: Please circle the appropriate numeric for the phrase that best describes the actions you observe of the learner groups’ performance from the recorded simulation, and include additional notes as appropriate.

### Critical Thinking Skills Rating Instrument

#### IDENTIFY the Problem and Set Priorities

**Example Prompting questions:** What is going on? What is the chief problem? What is the most important issue?

<table>
<thead>
<tr>
<th>Not able to identify chief problem(s) of the patient</th>
<th>Able to identify chief problem(s), but was unable to prioritize key issues.</th>
<th>Able to identify chief problem(s), and was able to prioritize key issues with substantial (≥ 2 times) prompting</th>
<th>Able to identify chief problem(s), and was able to prioritize key issues with minimal (&lt;2 times) prompting</th>
<th>Able to identify chief problem(s), and was able to prioritize key issues without prompting</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
</tr>
</tbody>
</table>

#### DETERMINE Relevant Information and Deepen Understanding

**Example Prompting questions:** How do the results relate to your patient? Do we all agree? Why do you think that?

<table>
<thead>
<tr>
<th>Not able to verbally express correct physiological concepts underlying patient’s condition</th>
<th>Able to verbally express some physiological concepts correctly and some incorrectly underlying patient’s condition</th>
<th>Able to verbally express physiological concepts underlying patient’s condition correctly with substantial prompting (≥ 2 times)</th>
<th>Able to verbally express physiological concepts underlying patient’s condition correctly with minimal prompting (&lt; 2 times)</th>
<th>Able to verbally express physiological concepts underlying patient’s condition correctly without prompting</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Not able to assimilate information from sources</th>
<th>Able to gather information from sources but not able to find relationships</th>
<th>Able to assimilate information from sources and find relationships with substantial prompting (≥ 2 times)</th>
<th>Able to assimilate information from sources and find relationships with minimal prompting (&lt; 2 times)</th>
<th>Able to assimilate information from sources and find relationships without prompting</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
</tr>
</tbody>
</table>

#### ENUMERATE Options and Anticipate Consequence

**Example Prompting Questions:** Why are you choosing this option? What are the options? What will occur as a result of your choice?

<table>
<thead>
<tr>
<th>Not able to discuss options for treatment</th>
<th>Able to discuss options for treatment, but was not able to anticipate consequence of treatments</th>
<th>Able to discuss options for treatment, and was able to anticipate consequences of treatments with substantial prompting (≥ 2 times)</th>
<th>Able to discuss options for treatment, and was able to anticipate consequences of treatments with minimal prompting (&lt; 2 times)</th>
<th>Able to discuss options for treatment and was able to anticipate consequences of treatments without prompting</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
</tr>
</tbody>
</table>

#### ASSESS the Situation and Make a Preliminary Decision

**Example Prompting Questions:** How did you come up with that conclusion?

<table>
<thead>
<tr>
<th>Not able to come up with diagnosis</th>
<th>Able to diagnose patient but was unable relate physical exam and lab findings with diagnosis</th>
<th>Able to diagnose patient and can relate physical exam and lab findings with diagnosis with substantial (≥ 2 times) prompting</th>
<th>Able to diagnose patient and can relate physical exam and lab findings with diagnosis with minimal promptings (&lt; 2 times)</th>
<th>Able to diagnose patient and can relate physical exam lab findings with diagnosis without prompting</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
</tr>
</tbody>
</table>

#### SCRUTINIZE the Process and Self-Correct as Needed

**Example Prompting Questions:** Did the patient improve or decline? Why?

<table>
<thead>
<tr>
<th>Not able to determine whether the patient was improving or declining</th>
<th>Was able to determine whether the patient was improving or declining but did not provide possible reasons</th>
<th>Was able to determine whether the patient was improving or declining and provided possible reasons with substantial prompting (≥ 2 times)</th>
<th>Was able to determine whether the patient was improving or declining and provided possible reasons with minimal prompting (&lt; 2 times)</th>
<th>Was able to determine whether the patient was improving or declining and provided possible reasons without prompting</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
</tr>
</tbody>
</table>

Comments:

---

Fig. 1. The critical thinking skills rating instrument.
able, with simulation 2 slightly outside the range (32). A third independent rater was asked to view and rate simulations 2 and 3, but the results were identical.

Correlations Between Knowledge and CTSRI Ratings

There is significant agreement that, for CT skills to develop, it is necessary to have a basic, foundational knowledge for the situation encountered. To determine whether there were relationships between knowledge (quiz performance) and CT skills (CTSRI video ratings), associations between knowledge and CT were performed for each of the six observations. Associations ranged from 0.04 to 0.35 for all four simulations. There were positive and significant correlations for the Determine (a) \([r_6 (89) = 0.26, P = 0.016]\), Enumerate \([r_6 (90) = 0.35, P = 0.001]\), and Assess \([r_6 (89) = 0.21, P = 0.049]\) observations. There were no significant correlations for the Identify, Determine (b), or Scrutinize observations. These data suggest that groups with a stronger knowledge base perform better in some CT skills, as defined in this study.

Median Ratings Between Groups First and Last Simulation

Next, median ratings for groups per observation were compared with Determine, if there was an improvement in CT skills of the whole class. There was a significant increase between the first and last HFPS in the median rating of groups for each of the six observations. The interquartile ranges were wider in the last HFPS compared with the first HFPS in five of the six observations, suggesting there was more variability in group CT at the end of the HFPS series. (Table 2).

Next, histograms comparing the frequency of a video rating for the first and last simulations were constructed to provide further insight into the data (Fig. 2). Ideally, as student groups develop better CT skills, they would rely less on faculty prompting, which would result in a video rater score of 3 or higher on the CTSRI. For example, in the Identify observation, ~27% of video ratings (8 out of 30) were 3 and above in the first simulation. In the last simulation, frequency of 3 and above ratings increased to 60% (18 out of 30). There was a rightward shift on all six observations as the percentage (%) of above ratings increased to 60% (18 out of 30). There was a rightward shift on all six observations as the percentage (%) of video ratings of 3 or higher increased between the first and last HFPS in the median rating of groups. There was a significant increase between the three other observations may be because these observations had the highest median on the first simulation (2.0) compared with the other three observations (Determine (a), Enumerate, and Assess), thereby limiting the spread of those data points. It is also possible that there were differences in the ways that the facilitator prompted each group. For example, one facilitator may let students discuss more before prompting, which could result in differences in video ratings. We tried to control for this by meeting as a group before each HFPS to go over the manikin program and standardizing each facilitator’s general approach to each HFPS. It is well known that standardization of any instructional delivery is a challenge and always needs to be considered.

Another key finding from this study was a significant increase in the median score for the class between the groups’ first and last simulations. Further exploration into this using histograms show that, despite the increase in overall median score, there are still many observations of groups needing substantial prompting (video rating of 2 or less). For example, for Determine (a), there were 15 observation of groups (50%) that needed substantial prompting in their last simulation; for Enumerate, there were 14 observations (47%) that indicated groups needed substantial prompting. One explanation could be dysfunctional group dynamics. De Grave et al. (6) have shown that lack of cohesion and motivation inhibited the learning process in a PBL class. It is possible that this occurred in some of the tested groups. If possible, groups included only one person with health professional experience. Therefore, some groups may have had members with previous experience, such as emergency medical technician training or nurse training, while other groups lacked such representation. Other group dynamics that potentially affected results were groups with one strong, loud personality where the rest of the group were more quiet individuals. This could be important to group dynamics, if the loud student is also overconfident, as this has been associated with diagnostic errors in medicine (3).

### DISCUSSION

The overall purpose of this pilot study was to determine the feasibility of observing group physiology CT skills in HFPS settings. To do this, an instrument, the CTSRI, was developed. Data from the current study suggest that the CTSRI is a valid and reliable instrument for rating CT skills, and that the ratings are valid for most of the tested HFPS scenarios. One finding from this experiment is the correlation between basic knowledge and CT skills in three of six observations. Therefore, we believe there is preliminary support for construct validity of the CTSRI, despite no correlations between basic knowledge and three of the observations [Identify, Determine (b), and Scrutinize]. This is crucial, since it is well established that knowledge underlies any CT process (9, 10, 15, 22). The lack of correlation between the three other observations may be because these observations had the highest median on the first simulation (2.0) compared with the other three observations [Determine (a), Enumerate, and Assess], thereby limiting the spread of those data points. It is also possible that there were differences in the ways that the facilitator prompted each group. For example, one facilitator may let students discuss more before prompting, which could result in differences in video ratings. We tried to control for this by meeting as a group before each HFPS to go over the manikin program and standardizing each facilitator’s general approach to each HFPS. It is well known that standardization of any instructional delivery is a challenge and always needs to be considered.

Another key finding from this study was a significant increase in the median score for the class between the groups’ first and last simulations. Further exploration into this using histograms show that, despite the increase in overall median score, there are still many observations of groups needing substantial prompting (video rating of 2 or less). For example, for Determine (a), there were 15 observation of groups (50%) that needed substantial prompting in their last simulation; for Enumerate, there were 14 observations (47%) that indicated groups needed substantial prompting. One explanation could be dysfunctional group dynamics. De Grave et al. (6) have shown that lack of cohesion and motivation inhibited the learning process in a PBL class. It is possible that this occurred in some of the tested groups. If possible, groups included only one person with health professional experience. Therefore, some groups may have had members with previous experience, such as emergency medical technician training or nurse training, while other groups lacked such representation. Other group dynamics that potentially affected results were groups with one strong, loud personality where the rest of the group were more quiet individuals. This could be important to group dynamics, if the loud student is also overconfident, as this has been associated with diagnostic errors in medicine (3). In our

### Table 2. Comparison of median ratings for each observation from all of the groups’ first and last simulation

<table>
<thead>
<tr>
<th>Observation</th>
<th>First Simulation</th>
<th>Last Simulation</th>
<th>P Value</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Median rating of groups</td>
<td>Interquartile range</td>
<td>Median rating of groups</td>
</tr>
<tr>
<td>1. IDENTIFY the problem and set priorities*</td>
<td>2.0</td>
<td>1.0–2.8</td>
<td>3.0</td>
</tr>
<tr>
<td>2. DETERMINE relevant information and deepen understanding (a)*</td>
<td>1.0</td>
<td>1.0–2.0</td>
<td>2.5</td>
</tr>
<tr>
<td>3. DETERMINE relevant information and deepen understanding (b)*</td>
<td>2.0</td>
<td>1.0–2.0</td>
<td>3.0</td>
</tr>
<tr>
<td>4. ENUMERATE options and anticipate consequence*</td>
<td>1.0</td>
<td>1.0–2.0</td>
<td>3.0</td>
</tr>
<tr>
<td>5. ASSESS the situation and make a preliminary decision*</td>
<td>0.5</td>
<td>0.0–2.0</td>
<td>3.0</td>
</tr>
<tr>
<td>6. SCRUTINIZE the process and self-correct as needed*</td>
<td>2.0</td>
<td>1.0–2.0</td>
<td>2.0</td>
</tr>
</tbody>
</table>

Values are median rating of all group video rating scores with the 25th and 75th interquartile range. *Significant change from first to last simulation; exact P value is provided.
HFPS, the second part of each debriefing session focuses on group dynamics to address some of these concerns, as past studies have shown that group dynamics can improve over series of HFPS, if these are included as a topic in debriefs (27). For this study, the initial discussion of teamwork skills was addressed after the first simulation. Therefore, as students improve teamwork skills, it is possible that this could also contribute to improved scores on the CTSRI (27). Further studies are still necessary to look at group interaction and the role it plays in the group thought process.

Another consideration is that the raters may expect an improvement across simulations and could bias the results. We did our best to control for this bias by providing coded videos to the raters so they were unaware of the sequence of simulations. One of the reasons the faculty were asked to participate as raters was that both were new to the institution and less familiar with the curriculum. Second, we recruited them because they did not have an active role in the module in which simulations took place. Although we did our best to control for rater bias and do not believe it occurred, it cannot be entirely ruled out.

Utility of the Instrument

The CTSRI has the potential to be an alternative assessment instrument as it is further developed. For example, the CTSRI could be used to identify groups in HFPS sessions that have insufficient physiology CT skills. Educators could intervene more rapidly by helping these groups identify their weakness and overcome their barriers. If faculty use the CTSRI in real time, the results could be shared directly with groups and be the focal point during the debriefing session. Medical students state that they rarely receive feedback, so the CTSRI may be a mechanism to provide it. Medical schools could also use the CTSRI to generate a narrative assessment of students to address Standard 9 of the Liaison Committee on Medical Education standards. In the future, the CTSRI may be an instrument for summative assessment of physiology during HFPS. To our knowledge, there are no studies on summative assessment and HFPS for medical students outside of clinical skills encounters.

Limitations of the Study

One limitation of the study is that it only measures group CT skills and not individuals. Therefore, it is unknown whether all members of the group possess improved CT skills. As mentioned previously, group dynamics could also play a role in the development of CT skills. Future research could use the IDEAS framework to capture individual data during a HFPS. For example, before a group can take on the task, each individual would have to respond independently to what he or she believes the chief problem is and why. This type of design could also allow the researcher to investigate confidence and its effect on group dynamics. While this may slow down the HFPS and alter its rhythm, it may be valuable for the learning process, especially for novice learners.

Another limitation of the study is that it is challenging for educators to determine the difficulty of a HFPS. This is hard to quantify and is dependent on a student’s strengths and weak-
nesses, but HFPS difficulty certainly could play a role in the groups’ thought processes. The challenge of HFPS for novice learners could also lead to cognitive overload, which is a valid criticism against the use of HFPS (12). Data from the present study suggest that, if students do not have the appropriate basic knowledge background, they may not be able to think critically. This supports the need to limit intrinsic cognitive load in novice learners. There are also heightened states of emotions in learners during HFPS that could influence the learning of some students. Still, HFPS are one of the most engaging modalities. Most of the research in simulation in medical education has focused on clinical skills or resident training, with very little focusing on novice learners, such as first-year medical students. Future investigations could explore the balance of engagement, cognitive overload, and the role of affect (emotions) in novice learners during HFPS.

One factor that needs consideration in the current study is that CT occurs over time, regardless of the HFPS teaching and assessment. A true control group (no HFPS) was not included because of the need for comparable experiences. Therefore, we cannot state that HFPS contributed solely to CT skill development, although it did provide a platform to observe these skills. Future research could determine whether the structure of instruction around HFPS plays a role in the development of CT skills.

Conclusions

A series of HFPS sessions provides a feasible platform to observe group CT skills by using a novel instrument constructed using the IDEAS framework. To our knowledge, the present study is the first to capture group CT skills in first-year medical students using HFPS and thus to demonstrate improvement between the first and last simulations measured by using the median video rating of all groups. Based on findings from this study, if one intent of HFPS is to develop physiology CT, we strongly recommend that first-year medical students demonstrate competency in the basic physiology knowledge required for the HFPS before entering the session. Students could complete a quiz or task on the necessary basic knowledge for entry into the HFPS session. If not, the risk of cognitive overload increases and the possibility of a poor learning experience is more likely to occur.

ACKNOWLEDGMENTS

The authors acknowledge the help of the staff and technicians of the Clinical Skills and Simulation Center, especially Carolina Marchena and Philip Hendricson, especially Carolina Marchena and Philip Hendricson.

DISCLOSURES

No conflicts of interest, financial or otherwise, are declared by the authors.

AUTHOR CONTRIBUTIONS


REFERENCES


