AWARD NUMBER: CDMRPL-16-0-DM160013

TITLE: **Developing Assessment Tools to Better Understand the Mechanisms of Clinical Reasoning in Military Medical Simulation**

PRINCIPAL INVESTIGATOR: Steven J. Durning; Anthony R. Artino, Jr. (co-PIs)

CONTRACTING ORGANIZATION: Henry Jackson Foundation for the Advancement of Military Medicine

REPORT DATE: Feb 2020

TYPE OF REPORT: Final Report

PREPARED FOR: U.S. Army Medical Research and Materiel Command
Fort Detrick, Maryland 21702-5012

REPORT DOCUMENTATION PAGE

Form Approved
OMB No. 0704-0188

1. REPORT DATE	**2. REPORT TYPE**	**3. DATES COVERED**
Feb 2020	Final	**10/06/2016 - 10/05/2019**

4. TITLE AND SUBTITLE

Developing Assessment Tools to Better Understand the Mechanisms of Clinical Reasoning in Military Medical Simulation

5a. CONTRACT NUMBER

5b. GRANT NUMBER
CDMRPL-16-0-DM160013

5c. PROGRAM ELEMENT NUMBER

6. AUTHOR(S)
Steven J. Durning, MD, PhD

5d. PROJECT NUMBER

5e. TASK NUMBER

5f. WORK UNIT NUMBER

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)

UNIFORMED SERVICES UNIVERSITY
Bethesda, MD 20814

8. PERFORMING ORGANIZATION REPORT NUMBER

9. SPONSORING / MONITORING AGENCY NAME(S) AND ADDRESS(ES)
U.S. Army Medical Research and Materiel Command
Fort Detrick, Maryland 21702-5012

10. SPONSOR/MONITOR'S ACRONYM(S)

11. SPONSOR/MONITOR'S REPORT NUMBER(S)

12. DISTRIBUTION / AVAILABILITY STATEMENT
Approved for Public Release; Distribution Unlimited

13. SUPPLEMENTARY NOTES

14. ABSTRACT
This project aims to better understand how physicians engage in clinical reasoning and how different types of contextual factors influence their clinical reasoning performance and self-regulation. In addition to asking participants about how they think about reasoning we will also examine several biometric indicators such as heart rate and blood pressure to examine how participants' physiologic responses might be influenced while they are engaging in clinical reasoning. Understanding how contextual factors influence clinical reasoning performance and its self-regulation will help inform new educational interventions to help faculty and physicians learn how to improve their clinical reasoning.

15. SUBJECT TERMS-

16. SECURITY CLASSIFICATION OF:			**17. LIMITATION OF ABSTRACT**	**18. NUMBER OF PAGES**	**19a. NAME OF RESPONSIBLE PERSON**
					USAMRMC
a. REPORT	**b. ABSTRACT**	**c. THIS PAGE**			**19b. TELEPHONE NUMBER** *(include area code)*
U	U	U	UU		

Table of Contents

INTRODUCTION

The detrimental effects of diagnostic error are well established, making it essential to a) understand physicians' clinical reasoning, b) what affects clinical reasoning and c) how to mitigate any potentially negative influences. Clinical reasoning—encompassing medical knowledge, clinical information and the specifics of the clinical situation (the context)—entails the steps up to and including arriving at a diagnosis and treatment plan for a patient. In some cases clinical reasoning performance is affected by context specificity, i.e. when a physician arrives at two different diagnoses for two different patients having the same symptoms, findings and, ultimately, diagnosis. In these cases, where factors other than the case content influence the clinical reasoning performance, there are c*ontextual factors* at play. These can be divided into a) patient factors (e.g., language barriers), b) physician factors (e.g., burnout, sleepiness), and c) environmental factors (e.g., faulty electronic health records). (See Figure 1.) This project sought to explore the influence of these contextual factors, examining how physicians engage in clinical reasoning. This grant was informed by three theories: situated cognition theory, cognitive load theory and self-regulation.

The three general aims of the study were:

1. To examine how theoretically derived variables are related to clinical reasoning performance in vitro (video condition)
2. To examine how theoretically derived variables are related to clinical reasoning performance in-vivo (live scenario-based condition).
3. To evaluate whether a novel intervention based on the results of specific Aims 1 and 2 improves clinical reasoning performance.

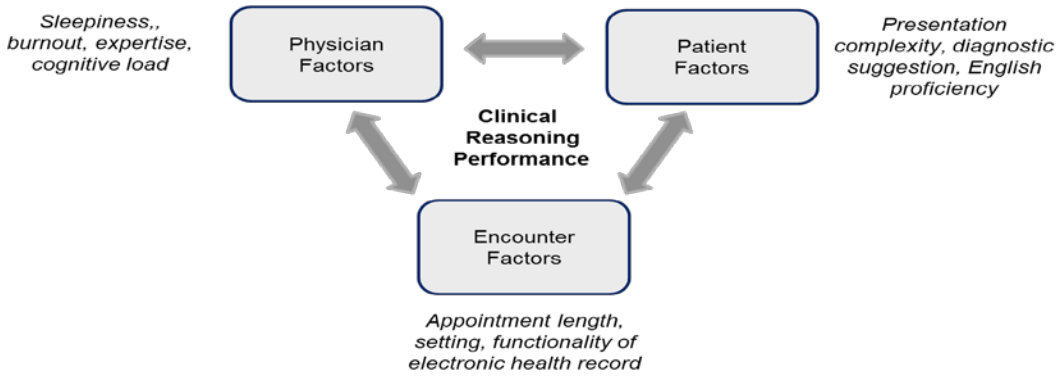
In order to examine how contextual factors influence physicians' reasoning performance, participants engaged in two different simulated patient encounters, a video condition (where participants watched two video-recorded clinical encounters with and without contextual factors;

- Aim 1) and a live scenario condition (where participants engaged in two clinical encounters with a standardized patient, one with and one without contextual factors;

- Aim 2) To measure the influence of different contextual factors, we examined: a) self-reported physician burnout, stress, sleepiness, and cognitive load, b) heart rate variability as a biometric marker of cognitive load, c) sleep patterns with an Actigraphy watch, d) physician self-regulation through microanalysis questions, e) clinical reasoning performance using an open-ended post encounter form and f) physician reflection on their clinical reasoning performance while they were engaged in a clinical encounter by asking them to 'think aloud' while either re-watching the video recorded clinical encounter (video condition) or watching their own recording of the clinical encounter (live scenario condition).

- Finally, using what we learned about how contextual factors influence clinical reasoning performance and self-regulation from Aims 1 and 2, we designed and tested a computer-based educational intervention (Aim 3) to help physicians learn how to potentially improve their clinical reasoning in the presence of contextual factors.

Figure 1: Situated Cognition as a Theoretical Framework for Context within a Clinical Encounter.



KEYWORDS

Clinical reasoning, Context specificity, Video-based simulation, Simulation-based learning, Situated cognition theory, Cognitive load theory, Self-regulated learning, Microanalysis, Biometrics, Think-aloud, Burnout, Sleepiness, Contextual factors

ACCOMPLISHMENTS

3.1) What were the major goals of the project?

The goals of the project are encapsulated in our aims. We divided the three general aims into three phases and sub-aims for each phase

Phase 1

- *Aim 1*: Video scenario case development and validation (completed)

  Milestone: completion of videos (completed more videos than stated in proposal)
- *Aim 2:* Live scenario case development and validation (completed)
  - *Milestone:* Scenario created (completed more live scenarios than stated in proposal)
  - *Deliverable 1:* Dissemination of case development process (completed)
- *Aim 3:* IRB submission and approval (completed)
  - *Milestone*: IRB Approvals (completed)

Phase 2

- *Aim 4:* Enrollment and data collection (both phases 2 and 3 completed)
- *Milestone:* Subjects enrolled and phase 2 and 3 data collection complete (exceeded minimum sample requirement of 80 participants for phase one; worked with 22 participants in phase 3).
- *Aim 5: D*ata analysis (Completed)
- *Milestone:* completion of data analysis for phase 2 and 3 (Completed)
- *Deliverable 2:* Dissemination of findings (we have presented at five conferences, have 6 manuscripts published (or accepted) in peer-reviewed journals we have 5 manuscripts under review and 5 manuscripts in process. Please see the dissemination of findings table under Aim 5, data analysis). We are also working on additional analyses and manuscripts beyond what we promised in the grant. These are currently in process and will be submitted as an addendum after their completion in approximately 6 months.

Phase 3

- *Aim 6:* Construction and piloting of educational intervention (completed)
- *Milestone*: educational intervention developed (completed) and IRB approved (completed)
- *Aim 7:* Educational intervention (completed; data collected with 22 participants)
- *Milestone*: completion of intervention pilot (completed) and accompanying manuscript (submitted)

3. 2) What was accomplished under these goals? (We list accomplishments across all report years here)

<u>For Aim 1 : Development and Validation of Videos</u>

Construction of Videotapes:

We developed and revised three video-based scenario scripts using a design-based participatory approach.

We also drew on the Jeffries Framework and the International Nursing Association for Clinical Simulation and Learning guidelines to design the workflow of these video scripts.

All video scripts were reviewed by physicians and simulation and standardized patient subject matter experts (SMEs) for clinical accuracy between November 2016 and February 2017.

We cast, and the standardized patient educator trained, six standardized patients (SPs).

We conducted several read-throughs of the video scripts. Read-throughs are organized reading around the table of the scenario script by the SPs and the director.

This step served to ensure that the scenario was clinically accurate and provided the SPs an opportunity to rehearse their roles to improve implementation fidelity.

We then filmed three non-contextual factors videos and three contextual factors videos (two videos for each of three conditions). We used four of the six videos for the study (exceeded the goal of construction of 4 videos).

Filming was completed on 1/31/2017.

Validation of Videos:

We conducted post-production editing of all six videos. To ensure fidelity of the videos we sought to ensure that the videos were similar in length and that the images of diagnostic findings (e.g., X-ray images, lab studies, EKG tracings) were visible for an appropriate and consistent length of time.

All scenarios were then re-reviewed by select physician SMEs and additional change requests were collected and informed additional revisions.

This process continued until we received no additional change requests.

We completed all post-production tasks on 3/23/2017.

All 6 videos were ready to use for the research study on 03/23/2017.

Milestone Achieved: IRB approval of use of videotapes with human subjects (December 2016)

For Aim 2: Live **Scenario Case Development and Validation**

Writing and revising outpatient and inpatient scenarios:

We developed and revised three live scenarios using a design-based participatory approach.

We adapted the Video 1 (Unstable Angina with contextual factors) and Video 2 (New Onset Diabetes without contextual factors) to become Live Scenarios 1 and 2 (Aim 2) to support study continuity [Attached as Supplemental Document].

We developed Scenario 3 (Aim 2) – Team-Based Trauma [Attached as Supplemental Document].

All scenarios were reviewed by physician, simulation and standardized patient subject matter experts (SME) for clinical accuracy between November 2016 and February 2017.

Validation of the live scenarios:

Multiple investigators conducted an analysis of the simulation contexts of the three live-scenarios using principles of activity theory. This served to ensure that each scenario accurately represented the clinical setting, that all necessary clinical artifacts and tools were noted and to identify potential extraneous cognitive load issues.

We cast and trained six standardized patients (SPs).

We conducted read-throughs of the scenarios, which are organized reading around the table of the scenario script by the standardized patients and select physicians.

This step served to ensure that the scenario was accurate and provided the standardized patients additional opportunities to rehearse their roles to improve implementation fidelity of the scenarios.

We conducted live rehearsals for the scenarios, which included the standardized patients portrayal of their role and interacting with the standardized patient trainer and select physicians who were not study participants (April 2017).

This step served to ensure that the scenario was accurate and provided the standardized patients additional opportunities to rehearse their roles to improve implementation fidelity of the scenarios.

Milestone Achieved: Completed implementation with 85 participants for General Aims 1 and 2 (Target of 70 participants; August 2018)

Deliverable 1: Dissemination of case development process

We developed and submitted a manuscript entitled, "Clinical Reasoning in the Primary Care Setting: Two Scenario-Based Simulations for Residents and Attendings" that was accepted for publication in *MedEdPortal* [Attached as Supplemental Document].

We developed and submitted a manuscript entitled, "Clinical Reasoning in the Medical Surgical Ward Setting: A Rapid Response Scenario for Residents and Attendings" that was accepted for publication in *MedEdPortal* [Attached as Supplemental Document]

For Aim 3: IRB Submission and Approval

IRB actions undertaken [2016-2020]:

Submitted initial application for IRB approval (approved -- 10/27/2016). [Attached as Supplemental Document]

Submitted modification adding Dr. Jeffrey LaRochelle as additional investigator [Attached as Supplemental Document].

Submitted modification for adding new research personnel Divya Ramani and Dr. Matt Ritter (approved - 08/17/2017). [Attached as Supplemental Document]

Submitted modification adding Dr. Abigail Konopasky, Sunny Yauger, and Dr. Luke Surry to the study protocol [Attached as Supplemental Document]

Submitted 2017 - 2018 continuing review to eIRB on 27th Sep 2018 (approved). [Attached as Supplemental Document]

Submitted modification adding Brooke Army Medical Center (approved) [Attached as Supplemental Document]

Submitted modification adding research personnel Megan Ohmer and Dr. Ajuzie, Dr. Haigney (approved) [Attached as Supplemental Document]

Submitted 2018-2019 continuing review on eIRB (approved) [Attached as Supplemental Document]

Submitted modification for adding new research personnel Dr. Jerusalem Merkebu (approved) [Attached as Supplemental Document]

Submitted modification for adding new research personnel Dr. Micheal Soh (approved) [Attached as Supplemental Document]

Milestone Achieved: Received IRB approval for latest continuing review 2019-2020. [Approved, 12/02/2019]

For Aim 4: Enrollment and Data Collection

For phase 2 of the study, we successfully met our target on 09/12/2018, enrolling and collecting data for **85** participants (attending and resident physicians).

For phase 3 of the study (intervention), we successfully completed the data collection with 22 participants (attending and resident physicians), from local and other sites [Brooke Army Medical Center (BAMC) and the University of Texas Health Science, San Antonio (UTHSCSA)]

Milestones Achieved: Completed data collection for video and live scenario conditions with 85 participants September 12, 2018. Data collection for intervention phase completed with 22 participants August 16th, 2019.

For Aim 5: Data Analysis

*The following table provides an overview of our diverse data analysis efforts for this project. We offer more detail on these efforts after this table. Please refer to dissemination of findings tables for information regarding publications.

Data Analysis	**As of 2020 February**
Think Aloud Transcription	Completed transcribing the think alouds of a total of all 107 participants
Task Based Coding (i.e., hand coding for clinical reasoning tasks followed by quantitative analysis in SPSS)	Completed data analysis (i.e., coding 60 think alouds).
Semantic Qualifiers (i.e., hand coding for level of clinical complexity of participant language followed by quantitative analysis in SPSS)	Completed data analysis by coding a sample of think alouds (*n* = 68).
Outpatient post-encounter form (PEF) (i.e., hand coding of clinical reasoning performance measure for outpatient scenarios followed by quantitative analysis in SPSS	Completed coding all outpatient PEFs for phases 2 and 3 (*n* = 85).
Inpatient PEF (i.e., same analysis as prior, but for inpatient scenario)	Completed coding all inpatient PEFs (*n* = 20).
Self Regulation Analysis (i.e., Rutgers team led by Dr. Timothy Cleary analyzed microanalytic questions for adaptive and maladaptive regulation responses)	Completed data analysis by coding all the outpatient data (n=64).
Actigraphy (i.e., quantitative analysis using SPSS of Actigraphy watch data on sleep patterns of participants)	Completed data analysis by coding all the Actigraphy watch data (n = 32)
Holter (i.e, analysis of biometric data by cardiology team using Mortara software followed by quantitative analysis using	Completed data analysis by coding all the holter data (n = 30)

SPSS)	
Linguistic Analysis (i.e., analysis of transcribed think aloud data using Linguistic Inquiry and Word Count)	Completed coding 40 think alouds (n = 25) analyzing linguistic markers to see the difference between video and live conditions.
Qualitative Analysis	We looked at physicians' think-aloud transcripts, exploring a) differences between video and live simulation b) physicians' reconsideration of choices, and c) linguistic markers of uncertainty.

Analyzing data is a two-step process which involves: 1) managing and transforming data and 2) conducting the analysis.

1. *Managing and transforming data*: We follow a data management system to ensure all data are retrieved and saved for analysis. The data saved is then transformed into a format on which we can run analysis, which involves:
    a. Retrieving and saving all video data to dual hard drives.
    b. Retrieving and saving all audio to dual hard drives
    c. Transcribing think-aloud interviews for a total 107 participants
    d. Transcribing live scenario patient encounters for a total of 20 participants
    e. Interpreting Mortara data from Holter monitor (done by cardiology team)
    f. Interpreting and coding microanalysis data (done by Rutgers self-regulated learning team)

2. *Analysis*: Since we have a diverse set of collected data (e.g. biologic data, self-report surveys, think alouds, videos, self-regulation data, etc.) and are working to create a series of different clinical reasoning assessment tools, we have called on different teams of expert researchers to analyze these data in different ways. Below is additional detail on some of this innovative analysis.
    a. Biometric data

i. *Holter analysis:* An electrophysiologic team analyzed Holter data of 32 participants looking for heart rate variability and QT variability. These offer information about the sympathetic and parasympathetic nervous systems which we then trace throughout our different (i.e., with and without contextual factors) encounter for each individual.

ii. *Actigraphy analysis*: With the help of a sleep specialist, we were able to interpret the sleep patterns of our participants who wore the Actigraphy watch for several days prior to the study (n = 32). We then sought correlations between these sleep patterns and clinical reasoning performance.

b. Think-alouds and other transcribed language data:

i. *Linguistic Inquiry and Word Count (LIWC) analysis:* Used a software for linguistic coding called LIWC to compare (a) transcripts of selected participants in live versus video encounters (n = 24) and (b) transcripts of each participant in cases with and without contextual factors (n = 64). This software allowed us to examine linguistic markers like pronouns, verb types, and words with affective significance. (See LIWC's website for more information: https://liwc.wpengine.com/)

ii. *Semantic competence analysis:* To investigate how cognitive load might have manifested in diminished expert performance, we created a coding scheme for semantic competence and discompetence, analyzing participants' think alouds for higher-level medical terminology (or instances where participants could have used higher-level medical terminology but didn't) as a marker of higher order thinking.

c. *Clinical reasoning task analysis:* With the help of two clinicians who helped to develop a published clinical reasoning task coding schema (see Supplemental Document A for this published schema), we examined patterns in different types of clinical reasoning tasks (e.g., diagnosis, reflection) with and without contextual factors.

d. *Survey data:*

    i. *Post Encounter Form (PEF) analysis: We assembled a team of three clinicians at USUHS who had done prior work developing the PEF, an open-ended, authentic measure of clinical reasoning performance with validity evidence. Each data point provided by participants (e.g., a suggested differential diagnosis or a potential lab test) was coded as correct, partially correct, or incorrect and then scores were developed for seven different aspects of clinical reasoning (e.g., future exam actions, leading diagnosis, evidence for a diagnosis). All participant PEFs were scored by this team (n = 170, due to most participants doing two cases). This was a primary outcome of several of our analyses.*

e. *Microanalytic data:*

    i. *Self-regulated learning analysis*: Based on prior published work on self-regulation, Dr. Timothy Cleary led a team of researchers at Rutgers university who coded our microanalytic questions (e.g., if you were to do this case again, what would you do differently?) to determine whether contextual factors or simulation environment (e.g., live versus video) affects the application of self-regulated learning strategies.

Deliverable 2: Dissemination of findings in Journals and Conference Presentations

This table offers details on the progress of our dissemination of findings (please see data analysis table above for information on data analysis progress related to planned dissemination). Note that the items marked with an asterisk below are part of an upcoming special issue of *Diagnosis* on clinical reasoning and contextual factors. This issue draws heavily from this JPC-funded project.

General Aims	Title	Journal/Conference	Status
Manuscript addressing Aim 1 of the study			
Aim 1	"The Linguistic Effects of Context Specificity: Exploring Affect, Cognitive Processes, and Agency in	*Diagnosis*	Published [Attached as Supplemental Document]

	Physicians' Think-Aloud Reflections"		
Aim 1	"Understanding Context Specificity: The Effect of Contextual Factors on Clinical Reasoning"	*Diagnosis*	Under Review [Attached as Supplemental Document]
Manuscripts addressing Aim 2 of the study			
Aim 2	"Clinical Reasoning in the Primary Care Setting: Two Scenario-Based Simulations for Residents and Attendings."	*MedEdPortal*	Published [Attached as Supplemental Document]
Aim 2	"Clinical reasoning in the inpatient setting: A standardized patient case for residents and attendings."	*MedEdPortal*	Published [Attached as Supplemental Document]
Aim 2	"Effects of Live and Video Simulation on Clinical Reasoning Performance and Reflection"	*Medical Education*	Addressing reviewers' comment [Attached as Supplemental Document]
Aim 2	"Sequence Matters: Patterns in Task-Based Clinical Reasoning" to Journal of Diagnosis	*Diagnosis*	Under Review [Attached as Supplemental Document]
Aim 2	"Examining Patterns of Uncertainty across Clinical Reasoning Tasks: The Effects of Contextual Factors on Clinical Reasoning Performance"	*Diagnosis*	Under Review [Attached as Supplemental Document]
Aim 2	"It Totally Possibly Could Be: How a Group of Military Physicians Reflect on Their Clinical Reasoning in The	*Military Medicine*	Accepted [Attached as Supplemental Document]

	Presence of Contextual Factors"		
Manuscript addressing Aim 3 of the study			
Aim 3	"Awareness and Reflection: The Results of an Intervention to Address Context Specificity"	*Diagnosis*	Under Review [Attached as Supplemental Document]
Additional manuscripts addressing more than one Aim			
	"First-Year Medical Students' Calibration Bias and Accuracy Across Clinical Reasoning Activities: An Initial Investigation"	*Advances in Health Sciences Education*	Accepted [Attached as Supplemental Document]
	"Why HPE Needs Functional Linguistics: The Power of Grammatical Categories to Medical	*Medical Education*	Published [Attached as Supplemental Document]
	"Five Principles for Using Educational Theory: Strategies for Advancing Health Professions Education Research".	*Academic Medicine*	Published [Attached as Supplemental Document]
	Heart Rate and Heart Rate Variability Correlate with Clinical Reasoning Performance and SelfReported Measures of Cognitive Load"	*Scientific Reports*	Published [Attached as Supplemental Document]
Manuscripts in process			
Aims 1 & 2	Manuscript based on Holter Monitor Data	*TBD*	Draft in process

Aim 2	Manuscript based on inpatient PEF Data	*TBD*	Draft in process
Aim 2	Manuscript on Semantic Qualifiers	*TBD*	Draft in process
Aim 2	Manuscript on qualitative analysis of think alouds (extremes of performance)	*TBD*	Draft in process
Aims 1 & 2	Well-Being in a Cohort of Active Duty Military Physicians: An Assessment of Sleep Patterns, Burnout and Perceived Stress	*MedEd Publish*	[Attached as Supplemental Document]
Aim 1 & 2	Qualitative Analysis of think alouds	*TBD*	Draft in process
Aim 1 & 2	In depth analysis of Self-regulation	*TBD*	Draft in process
Presentations			
Aim 1	"The Effect of Contextual Factors on Clinical Reasoning: A Mixed Methods Study Examining Outcome and Process."	*Presentation*	Accepted [Attached as Supplemental Document]
Aim 1	"Case Specificity in Clinical Reasoning: A Qualitative Case Study of Conditional Reasoning Processes"	*Poster Presentation*	Accepted [Attached as Supplemental Document]
Aim 2	"The Use of Think-Aloud Reflections to Examine Learners Experiences in Live and Video-Based Simulation Contexts: A Comparison	*Presentation*	Accepted [Attached as Supplemental Document]

	Stud".		
Aim 2	"An Introductory Workshop for Activity and Linguistic Analysis of Video in Healthcare Simulation"	*Workshop*	Accepted [Attached as Supplemental Document]
Aim 2	"Emergent Clinical Reasoning During Think-Alouds: How Physicians Reflect on their Own and Others' Practices in Live and Video Simulation"	*Presentation*	Accepted [Attached as Supplemental Document]
Aim 2	"Examining the Influence of Simulation Context on Learners' Post-Simulation Reflections: A Comparison Study using Think Alouds".	*Presentation*	Accepted [Attached as Supplemental Document]
Aim 1	"Uncovering patterns of uncertainty across clinical reasoning tasks"	*Poster Presentation*	Under Review [Attached as Supplemental Document]

For Aim 6: Construction and Piloting of Educational Intervention

We completed a literature review, interviewed a clinical reasoning and diagnostic error expert, and used our empirical study data to develop our intervention [Attached as Supplemental Document]

We completed, reviewed and edited the intervention design and began data collection in March 2019 [Attached as Supplemental Document]

The intervention included:

> An interactive video training to teach intervention participants what factors affect clinical reasoning (contextual factors), how contextual factors can lead to diagnostic error, and reflection strategies to help overcome and improve clinical reasoning. At various stages of the training module we incorporate metacognitive

monitoring by asking participants to describe and reflect on experiences that correspond to what they are learning.

A think aloud reflection tool (based on our study instrument, but adapted for participants to use themselves) to give the participants [Attached as Supplemental Document]

The training curriculum was piloted with 2 physicians; we adapted their feedback and made modifications to the training videos as well as Qualtrics survey layout.

We also developed a new outpatient scenario case (Gallstone Pancreatitis) [Attached as Supplemental Document].

In order to ensure the reliability and authenticity of the developed case, we ran multiple rehearsals with standardized patients, physicians, and sim center personnel While we did not use this case in our final intervention design, we are including it as a product of grant funding.

Milestones Achieved: Successfully developed different components of the intervention phase: a) training video b) think aloud reflection and c) outpatient scenario case

<u>For Aim 7: Educational Intervention</u>

- We completed implementing the intervention [August 2019]. We collected data from three different study sites, Walter Reed Medical Center, Brooke Army Medical Center, and University of Texas Health Science and successfully gathered data from 22 participants.
- The initial intervention manuscript (based on PEF data) is completed and has been submitted to *Diagnosis*. We are currently in the process of analyzing the think-aloud transcripts, comparing the intervention and control condition for linguistic markers.

Milestone Achieved: Completed data collection with 22 participants

Additional tasks achieved:

In addition to the goals, tasks, subtasks and milestones achieved and reflected on above, we also achieved several additional infrastructure-related tasks, mitigating difficulties as they emerged, including:

Technical, eIRB, and Recruitment:

Technical:

We mitigated the technical difficulties i.e. internet disruption problems by ordering 2 MI-Fi. We also coordinated with the simulation center staff to provide us with backup laptops in the case of MI-Fi disruption.

We researched different software packages and gained approval from the USUHS IT department to purchase software packages for data analysis.

We contracted with a transcription company, Accentance Inc., to help us with transcribing audio recordings of think alouds and live scenarios.

eIRB:

Due to major delays in the eIRB process, we began working in conjunction with an eIRB official to help us resolve and address our questions so that we could get the modifications approved in a timely manner.

Recruitment:

Due to difficulties in recruitment, we offered participants flexible scheduling, both regarding time and location.

We also began using a "snowball" strategy, reaching out to prior participants for suggestions of other potential participants to recruit.

We scheduled recruitment sessions during the intern orientation (beginning in June) and faculty onboarding (in July) when participant schedules were more predictable.

Owing to recruitment challenges at local sites, we added new study sites to our study protocol via IRB to help us in recruiting participants for the study, which included NMCSD, BAMC, and UTHSCSA.

METHODOLOGY

Sample:

Phase 2 (comparative study): 85 primary care (internal medicine and family medicine) and surgical residents and attending physicians. (Our target was 80 participants based on a power analysis.)

Phase 3 (pilot intervention): 22 internal medicine and family medicine and surgical residents and attending physician). (Our target was slightly higher (n = 40), but based on preliminary analysis of the data we were able to gain meaningful information in terms of the effect of training on clinical reasoning (i.e. physicians' ability to overcome and mitigate the hindering effects of contextual factors).

Overall Design:

Mixed-methods, experimental design. Use of qualitative and quantitative measures is complementary.

Using video- and live scenario-based simulations to elicit physicians' reasoning in increasingly complex and authentic settings.

Comparison of video-and live-scenario-based simulations.

Adapting novel measures (e.g., biometric, think-aloud, self-regulated learning, linguistic) to assess physicians' reasoning.

Design and development of video- and live-scenario-based simulations.

Outpatient Videos and Scenarios & Inpatient Scenario :

> We developed, revised, and implemented 4 of 6 outpatient videos and two outpatient live simulations (all either diabetes or angina cases; completed Jan 2017) and one inpatient scenario (tension pneumothorax case; completed, May 2018) [Attached as Supplemental Document] [Please refer to section 3.2 of this report for further details].

Intervention Training Curriculum for Phase 3:

- Based on the results from phase 2 of the study, we developed a computer-based intervention training. We incorporated both expert opinion and recent literature: (a) We interviewed an expert in clinical reasoning and diagnostic error, with a primary goal of getting an expert opinion on strategies that can help physicians reduce diagnostic error. (b) We conducted an extensive literature review, focusing on new findings that have emerged since we began our study. (c) Finally, we synthesized the results of our study. We used those insights from all three of these processes to design a training curriculum.

- Next, one of our research assistants adapted the developed training curriculum into three separate sets of training videos. This process involved: a) developing a PowerPoint presentation; b) embedding animation; c) recording, editing and embedding voice-over (two individuals) for video; and d) uploading the training video that we embedded into our Qualtrics survey. [Attached as Supplemental Document]

- Next, we developed two surveys in Qualtrics for our intervention group and pilot tested the training with physicians on the team to assess the curriculum's reliability and validity. We adapted based on their feedback and made modifications to our module and surveys as necessary

- We also developed, rehearsed, and validated a new outpatient scenario case for the intervention (Gallstone Pancreatitis), but the design team determined it would be more efficacious to compare intervention participants with existing participant data, so we did not use this case, but include it as part of the deliverables for the grant [Attached as Supplemental Document].

- We began data collection for phase 3 in March 2019, and successfully completed phase 3 with 22 participants [August 2019].

3.3) What opportunities for training and professional development has the project provided?

- The project involved several integrated processes of data collection and management at different points in the study. We also developed an educational curriculum including training videos for Phase 3 (the intervention phase) of the study. We will be providing these videos along with other training materials (i.e. think aloud instructions), the 6 video cases and 3 live scenario cases.

- To ensure we collected usable data, followed IRB protocols, and appropriately secured the data, the key personnel and the research assistants were provided the following training opportunities:

Research Protocol Implementation Training: The study implementation entails a series of steps, for which each team member underwent training. The following are some of the training areas:

- sending initial recruitment emails to potential participants and corresponding with participants to schedule study times
- consenting participants
- setting up the live simulation rooms including the GoPro video system
- administering the pre- and post-study survey through Qualtrics on researcher computers (so that no identifiable data from participants was collected)
- conducting post-study think aloud reflections and microanalysis protocols
- removing data from digital devices and following designated data management protocols (discussed in more detail below)

1. **Proper Handling and Use of Biometric Devices:**

Actigraph watch: Members of the team were trained to configure the actigraphy watch using the actigraphy software and also on how to retrieve and save the data collected.

Holter monitor: Members of the team learned how to fit a 12-lead Holter monitor which involves 1) instructing the participant on how to wear the Holter monitor, 2) prepping the participant prior to fitting them with 12-lead Holter, b) fitting the Holter monitor, and d) retrieving the data and cleaning the Holter card.

Automatic blood pressure machine: Training was provided so that research team members could efficiently use the blood pressure machine and accurately record results.

AURiS stethoscope: Members of the study team learned how to use the Auris software to remotely create different breath sounds to represent worsening of the pneumothorax and deterioration of the patient. They also learned how to orient participants regarding the use of the stethoscope.

2. **Qualitative and Quantitative Data Analysis Training**:
   - *Qualitative think aloud analysis*: Team members assistants learned how to qualitatively code participants' think alouds, identifying: (a) instances of reconsideration: moments when practitioners questioned their own choices or thought processes and (b) uncertainty markers like *maybe* and *could*.
   - *Qualitative semantic competence*: One study team member (Megan Ohmer) was trained by the PIs to identify advanced medical terminology (semantic competence) as well as instances where participants could have used medical terminology and did not (semantic discompetence).
   - *Quantitative data analysis*: One study team member (Divya Ramani) was trained in more advanced uses of SPSS statistical analysis software for multiple data analysis efforts.

3. **Implementation Evaluation Training**: A checklist was developed to ensure reliable implementation of the scenario cases in terms of actors' consistent portrayal of standardized patients. Members of the team received training in use of this implementation evaluation checklist to determine usability of live scenario videos

4. **Training of Standardized Patients**: With assistance from the simulation lab staff, team members were taught how to train and educate actors in portraying standardized patients (e.g., appropriate portrayal of symptomatology, responses to the questions posed by the study participants, knowledge about medical history/background). This involved team members offering multiple practice sessions with the actors prior to the initial simulation as well as refresher training occasionally throughout the study.

5. **Data Management Training**: One of our research assistants (Divya Ramani) completed an online course on "Research Data Management and Sharing" to better aid in managing the diverse data collected for the study. The courses covered the following areas a) understanding data, b) developing a data management plan, c) working with ongoing data collection, d) sharing data, and e) archiving data for future use. The course also highlighted some of common problems faced due to poor data management (which can lead to data loss) along with ways of mitigating these problems. Ms. Ramani revised some of our data management procedures as a result of this training.

6. **Regulatory Training:** As per IRB regulations, all research personnel completed research-specific regulatory training.

3.4) How were the results disseminated to communities of interest?

We have submitted manuscripts for publication to a number of leading journals (two of them appeared in *MedEdPortal*, which offers open access educational resources to health professionals). We have also presented at conferences, both local and international, and also plan for future conference presentations. Please refer to section 3.2 (Dissemination of findings) for further details on dissemination.

3.5) What do you plan to do during the next reporting period to accomplish the goals?

We will continue to analyze data with new lenses and write additional manuscripts for publication. Below are some of the components we will be focusing on:

Data Analysis:

Continue with the activity analysis coding of live scenario-based simulations (inpatient & outpatient cases) video recordings.

Run further analysis on Holter data

Manuscript Publication: Complete and publish the following papers

Paper on clinical reasoning activity analysis

Paper on Holter monitor results

Paper on self-regulated learning and activity systems

IMPACT

Describe distinctive contributions, major accomplishments, innovations, successes, or any change in practice or behavior that has come about as a result of the project relative to:

4.1) What was the impact on the development of the principal discipline(s) of the project?

We conducted qualitative and quantitative analysis on the data collected with the aim of understanding physicians' clinical reasoning processes and outcomes and determining what, if any, differences emerge between video versus live simulation conditions using theoretically grounded measures and methods. The following are examples of impact we expect our work will have on medicine's understanding, instruction, and remediation of clinical reasoning:

Aim 1: Our results suggest that the study of and intervention into the effect of *context* on physician reasoning performance are important for understanding context specificity and reducing *errors* and that this work should be continued and built upon.

Aim 1: The findings also suggest new tools for *measuring* physician reasoning, including aspects of physician language and biometric data reflect clinical reasoning performance and are being affected by contextual factors.

Aim 1: In addition to the importance of context, our study indicates that the *content* of the clinical case (i.e., what type of medical problem) can affect physician reasoning as well (and is consistent with prior work in clinical reasoning).

Aim 1: We also found that clinical reasoning differs according to level of *experience*, suggesting we might need to teach clinical reasoning differently depending upon how much experience physicians have.

Aim 2: Our results also indicate that performance as well as learning opportunities in *live scenario cases* versus *video cases* are different. Participants seem to use different kinds of cognitive processing and reflection in each modality, perhaps thinking and reflecting a bit more *deeply* in the live scenario cases. Further study of the implications of and use of different genres of simulation (i.e., video, live) could be important based on the findings from our preliminary analyses.

Aim 2: Our results shed additional light on the differential effects of simulation approach and the benefits of using context-specific assessment tools to uncover underlying cognitive judgments and reactions of medical professionals during clinical tasks.

Aim 2: We also explored possible sequential patterns of reasoning (task based reasoning), i.e. physicians' engagement in different categories of tasks looking at whether there were any difference in terms of presence and absence of contextual factors. Our preliminary qualitative results suggest that contextual factors affect what participants are *uncertain* about across different tasks. Future work in this vein could better support physicians in the critical area of diagnostic uncertainty.

Aim 3: Our brief intervention may impact clinical reasoning performance and mitigate some of the detrimental impact of context specificity.

Please review appendices of submitted and published papers for more details on impact.

4.2) What was the impact on other disciplines?

While we studied context specificity in physicians, we expect that this phenomenon occurs with other health professionals who care for patients and our findings could help inform assessment, remediation, and teaching on this topic in other health care disciplines.

4.3) What was the impact on technology transfer?

Nothing to report.

4.4) What was the impact on society beyond science and technology?

Diagnostic error is an international health care crisis, accounting for approximately 10% of patient deaths and hospital adverse events in the United States. The results of this study are directly applicable to a common cause of diagnostic error, context specificity (see introduction for definition and discussion). The papers published under the auspices of this grant provide vital and novel information about context specificity, how it can affect physicians at varying levels of experience, how it is experienced differently across training contexts (i.e., video versus live simulation), how one can assess context specificity, and what may be done to mitigate its effects. Nothing to report.

CHANGES/PROBLEMS

5.1) Changes in approach and reasons for change?

There have been no specific changes made since the date of our last report.

5.2) Actual or anticipated problems or delays and actions or plans to resolve them?

There were difficulties in terms of recruiting participants due to scheduling issues. We developed, revised and improved our recruitment strategies in an ongoing manner and were able to meet or exceed our recruitment targets. (Please refer to section 3.2, Additional task achieved)

5.3) Changes that had a significant impact on expenditures?

Not applicable.

5.4) Significant changes in use or care of human subjects, vertebrate animals, biohazards, and/or select agents?

Not applicable.

PRODUCTS

6.1) Publications, conference papers, and presentations*:*

See sections 6.2 - 6.8 below.

6.2) Journal publications

Following are all the lists of manuscripts published. [* We have 5 manuscripts that are under review, and 5 papers in draft stage and 5 manuscripts based on results from this study. Please refer to the table in section: Dissemination of findings for further details]

Konopasky A, Ramani D, Ohmer M, Battista A, Artino A, McBee E, Ratcliff T, Durning S. It Totally Possibly Could Be: How A Group Of Military Physicians Reflect On Their Clinical Reasoning In The Presence Of Contextual Factors. *Military Medicine* (In Press)

Konopasky A, Ramani D, Ohmer M, Durning S, Artino A, Battista A. Why Health Professions Education Needs Functional Linguistics: The Power of "Stealth Words" for Advancing Research and Education. *Medical Education*. (Published)

- Battista A, Konopasky A, Ramani D, Ohmer M, Mikita J, Howle A, Krajnik S, Torre D, Durning S. Clinical Reasoning in the Primary Care Setting: Two Scenario-Based Simulations for Residents and Attendings. *MedEdPORTAL*.14:10773. 2018.

- Ohmer M, Konopasky A, Durning, S, Ramani D, Nealeigh, M, Kucera W, Ordway S, Mellor T, Mikita J, Battista A. Clinical reasoning in the inpatient setting: A standardized patient case for residents and attendings. *MedEdPORTAL.* (In Press)

  Awarded *Editor's Choice* distinction as an exemplary resource.

The current JPC study results informed the publication of a paper in the *Journal of the American Medical Association*, entitled, "Management reasoning: Beyond the diagnosis" (See open access link: https://jamanetwork.com/journals/jama/fullarticle/2681495)
The current JPC projects Holter component was informed based on results from previous studies holter data. The manuscript entitled, "Heart Rate and Heart Rate Variability Correlate with Clinical Reasoning Performance and Self-Reported Measures of Cognitive Load"

6.3) Books or other non-periodical, one-time publications.
"Nothing to Report"

6.4) Other publications, conference papers, and presentations.

International Presentations:

Konopasky A. Battista A. Examining the Influence of Simulation Context on Learners' Post-Simulation Reflections: A Comparison Study using Think Alouds. Accepted Research Abstract to be presented as a poster at AMEE 2019, August 24 -28, 2019. Vienna, Austria.
Konopasky A, Battista A. The Use of Think-Aloud Reflections to Examine Learners Experiences in Live and Video-Based Simulation Contexts: A Comparison Study. Research Abstract presented at the 19th International Meeting on Simulation in Healthcare, January 27 - 30, 2019. San Antonio, TX.
Battista A, Konopasky A. An Introductory Workshop for Activity and Linguistic Analysis of Video in Healthcare Simulation. Workshop offered at the 19th International Meeting on Simulation in Healthcare, January 27 - 30, 2019. San Antonio, TX.

National Presentations:

Konopasky A, Battista A, Ramani D, Artino A, Durning S**.** Emergent Clinical Reasoning During Think-Alouds: How Physicians Reflect on their Own and Others' Practices in Live and Video Simulation. Accepted podium presentation at Learn Serve Lead 2018: The AAMC Annual Meeting; 2018 November 2 - November 6; San Antonio, TX.
Konopasky A, Battista A**,** Ramani D, Artino A, Durning S. The Effect of Contextual Factors on

Clinical Reasoning: A Mixed Methods Study Examining Outcome and Process. Accepted oral presentation at the Military Health System Research Symposium (MHSRS) 2018: 2018 August 20-23; Kissimmee, FL.

Local Presentation:

Ramani D, Konopasky A, Battista A, Artino A, Durning S.: Case Specificity in Clinical Reasoning: A Qualitative Case Study of Conditional Reasoning Processes. USUHS research day; 2018 May 29th; Bethesda, MD.

6.5) Website(s) or other Internet site(s)

Nothing to report.

6.6) Technologies or techniques

Nothing to report.

6.7) Inventions, patent applications, and/or licenses

Nothing to report.

6.8) ***Other Products***

Training:

We developed training video as part of the intervention phase of the study, by incorporating both expert opinion and recent literature into our training curriculum. We will be providing the videos to the funders on a separate CD.

Live scenarios (4 total):

Ohmer M, Durning S, Kucera W, Nealeigh W, Mellor T, Ordway S, Mikita J, Howle A, Krajnick S Battista A, Konopasky A, Ramani D, Battista A. Clinical reasoning in the inpatient setting: A standardized patient case for residents and attendings. 2018

Battista A, Konopasky A, Ramani D, Ohmer M, Mikita J, Howle A, Krajnick S, Torre D, Durning S. Clinical reasoning in the primary care setting: Two standardized patient cases for residents and attendings. 2017

We also developed, rehearsed, and validated a new outpatient scenario case for the intervention (Gallstone Pancreatitis)

Videos (6 total)

Battista A, Hemmer P, McBee E, Ratcliffe T, LaRochelle J, Howle A, Durning S. Clinical Reasoning in the Primary Care Setting: Two video-based clinical vignettes. 2016

Design and post-production for the adaptation of two control video-based scenarios for the JPC-1, CDMRP grant, Award # NH83382416

PARTICIPANTS & OTHER COLLABORATING ORGANIZATIONS

7.1) What individuals have worked on this project?

Name:	Steven J. Durning
Project Role:	*Principal Investigator*
Duration on project:	Jan 2016 - Present
Percent effort:	10%
Contribution to Project:	Supervision of all study personnel; direction of research design and study implementation; direction of data analysis and dissemination; helped to write all papers and assisted with all conference presentations

Name:	Anthony R. Artino Jr.
Project Role:	*Co- Principal Investigator*
Duration on project:	Jan 2016 - Present
Percent effort:	5%
Contribution to Project:	Collaborated with PI on direction of research design and study implementation; direction of data analysis and dissemination; helped to write papers and conference presentations

Name:	Alexis Battista
Project Role:	*Key Personnel*
Duration on project:	Jan 2016 - Present
Percent effort:	20%
Contribution to Project:	Helped direct study design and implementation; assisted in recruitment and data collection; helped direct data analysis and dissemination; lead instructional designer of video and live scenario-based simulations; supervised construction and validation of video and live-scenario-based simulations; developed data management plan; helped to write papers and conference presentations

Name:	Abigail Konopasky
Project Role:	*Key Personnel*
Duration on project:	2nd Oct 2017 - present
Percent effort:	100%

Contribution to Project:	Assisted in recruitment and data collection; helped direct data analysis and dissemination; helped to write papers and conference presentations

Name:	Divya Ramani
Project Role:	*Key Personnel*
Duration on project:	20th March 2017 - present
Percent effort:	100%
Contribution to Project:	Directed recruitment and outreach efforts including: establishing recruitment relationships and strategies, scheduling and coordinating study participants, direction and oversight of simulation lab scheduling and setup to established study quality standards. Oversaw Simulation lab coordination with the simulation lab operations staff, data collection (including think-aloud protocol), data management (e.g., capture, archival, data transformation, data auditing); engaged in data analysis and dissemination (e.g., drafting of literature reviews and manuscripts).

7.2) ***Has there been a change in the active other support of the PD/PI(s) or senior/key personnel since the last reporting period?***

Nothing to report.

7.3) ***What other organizations were involved as partners?***

Provide the following information for each partnership:

Organization Name: Walter Reed National Military Medical Center

Location of Organization: Bethesda, Maryland, 20814.

Partner's contribution to the project: Research site for the study

Financial support - "Not Applicable"

In-kind support - "Not Applicable"

Facilities *(e.g., project staff use the partner's facilities for project activities)*; Walter Reed provides their simulation laboratory space for study implementation

Collaboration *(e.g., partner's staff work with project staff on the project)*; Walter Reed provides simulation laboratory staff for help with project when needed

Personnel exchanges *(e.g., project staff and/or partner's staff use each other's facilities, work at each other's site)*; "Not Applicable"

Other. "Not Applicable"

Organization Name: Rutgers University

Location of Organization: New Brunswick, New Jersey

Partner's contribution to the project: Consulting related to the self-regulated microanalysis protocol (Dr Tim Cleary)

Financial support - $124,698.00 (Y1 $37,811.00)

In-kind support - "Not Applicable"

Facilities *(e.g., project staff use the partner's facilities for project activities)*; Not Applicable

Collaboration *(e.g., partner's staff work with project staff on the project)*; Project partner, Dr. Tim Cleary, provided guidance in developing the self-regulated learning microanalysis protocol and supported refining the fuller research protocol. Dr. Cleary is also engaged in supporting data analysis related to the microanalytic protocol.

Personnel exchanges *(e.g., project staff and/or partner's staff use each other's facilities, work at each other's site)*; "Not Applicable"

Other. "Not Applicable"

SPECIAL REPORTING REQUIREMENTS

Please find attached Quad Chart

APPENDICES

Please refer to Supplemental Documents

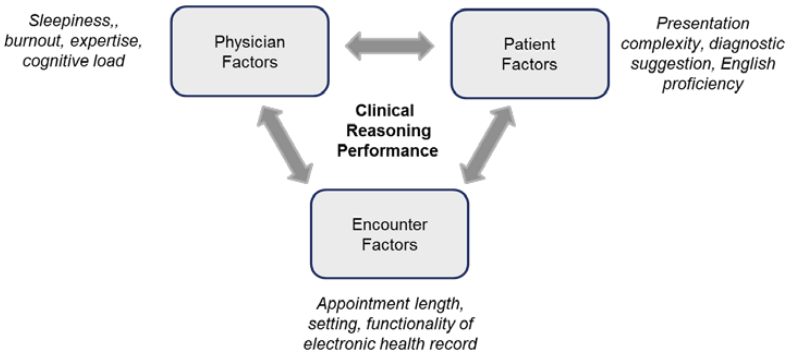
Quad Chart

Study Aims

•To examine how theoretically derived variables are related to clinical reasoning performance in-vitro using videotapes.
•To examine how theoretically derived variables are related to clinical reasoning performance in-vivo during live scenario-based simulations.
•To evaluate whether a novel intervention based on the results of specific aims 1 and 2 improves clinical reasoning performance.

Approach

This was a 3-phase, prospective, mixed-methods study design, which involved: (1) developing video recorded clinical encounters and scenario-based simulations; (2) using video tapes and a live, team-based trauma simulation scenario to investigate relationships between clinical reasoning performance, cognitive load and contextual factors; and (3) developing an intervention and studying its effects.



Timeline and Cost

Activities CY	16	17	18	19	20
Phase 1: Video development & validation					
Phase 2: Empirical studies (video and live)					
Phase 3: Intervention study					

Goals/Milestones:
CY20 Goal - Finish analysis and manuscript development
Comments/Challenges/Issues/Concerns:
☐ Recruiting for the intervention phase of the study
☐ **Budget Expenditure to Date:**
- Projected Expenditure: 1,397,500.05 $
- Actual Expenditure: 1,086475.4100 $

REPORT DOCUMENTATION PAGE		*Form Approved OMB No. 0704-0188*
Public reporting burden for this collection of information is estimated to average 1 hour per response, including the time for reviewing instructions, searching existing data sources, gathering and maintaining the data needed, and completing and reviewing this collection of information. Send comments regarding this burden estimate or any other aspect of this collection of information, including suggestions for reducing this burden to the Department of Defense, Washington Headquarters Services, Directorate for Information Operations and Reports (0704-0188), 1215 Jefferson Davis Highway, Suite 1204, Arlington, VA 22202-4302. Respondents should be aware that notwithstanding any other provision of law, no person shall be subject to any penalty for failing to comply with a collection of information if it does not display a currently valid OMB control number. PLEASE DO NOT RETURN YOUR FORM TO THE ABOVE ADDRESS.		
1. REPORT DATE 02/02/2020	2. REPORT TYPE Final Report	3. DATES COVERED 16/9/2016 - 2nd Feb 2020
4. TITLE AND SUBTITLE Developing Assessment tools to better understand the mechanisms of clinical reasoning in military medical simulation.		5a. CONTRACT NUMBER
		5b. GRANT NUMBER 308874-01.00-65126
		5c. PROGRAM ELEMENT NUMBER
6. AUTHOR(S) Report Authors: Dr. Steven Durning Dr. Anthony Artino Dr. Alexis Battista Dr. Abigail Konopasky Diya Ramani Associate Investigators: Dr. Steve Durning Dr. Anthony Artino Dr. Alexis Battista Dr. Abigail Konopasky Divya Ramani Megan Ohmer Dr. Timothy Cleary Dr. Paul Hemmer		5d. PROJECT NUMBER 308874

Dr. Elexis Mcbee Dr. Temple Ratcliffe Dr.Jeffrey Mikita Dr. Luke Surry Dr. Stephanie Ajuzie Dr. Mark Haigney Dr. Jerusalem Merkebu Dr. Micheal Soh E-Mail:	5e. TASK NUMBER 5f. WORK UNIT NUMBER
7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES) AND ADDRESS(ES) Uniformed Services University of Health Science 4301 Jones Bridge Road Bethesda, MD, 20814	8. PERFORMING ORGANIZATION REPORT NUMBER
9. SPONSORING / MONITORING AGENCY NAME(S) AND ADDRESS(ES) U.S. Army Medical Research and Materiel Command Fort Detrick, Maryland 21702-5012	10. SPONSOR/MONITOR'S ACRONYM(S) 11. SPONSOR/MONITOR'S REPORT NUMBER(S)

12. DISTRIBUTION / AVAILABILITY STATEMENT

Approved for Public Release; Distribution Unlimited

13. SUPPLEMENTARY NOTES

14. ABSTRACT

We explored the phenomenon of context specificity in clinical reasoning performance. We ran a mixed methods study, wherein 65 participants were quasi-randomly assigned to either outpatient video or outpatient scenario condition. We conducted both qualitative and quantitative analysis on the data we had collected, to examine clinical reasoning performance in the presence of contextual factors. We also examined the difference between the use of video versus live scenario conditions in terms of participant clinical reasoning performance. The findings revealed a significant difference between video versus live scenario condition in terms of participant reconsiderations and linguistic markers thereby signifying the effect of context on performance. Meanwhile comparing post encounter form (PEF) scores, we found a statistically significant difference in several items between contextual and non-contextual factor conditions. The presence of one or more contextual factors significantly and negatively impacted clinical reasoning as measured by the PEF.

The study also involves 20 participants assigned to an inpatient scenario condition. We ran preliminary content analysis on the think aloud data collected. The analysis revealed an association between participants' expertise level and their stated leading diagnosis.

Based on the results, we developed an intervention and successfully ran it with 22 participants. We are currently analyzing the data collected from the intervention phase.

15. SUBJECT TERMS

Clinical reasoning, Context specificity, Video-based simulation, Simulation-based learning, Cognitive load theory, Self-regulated learning microanalysis, Biometrics, Think-aloud, Burnout, Sleepiness, Situated cognition, Contextual factors

16. SECURITY CLASSIFICATION OF:			17. LIMITATION OF ABSTRACT	18. NUMBER OF PAGES	19a. NAME OF RESPONSIBLE PERSON
					USAMRMC
a. REPORT	b. ABSTRACT	c. THIS PAGE	Unclassified		19b. TELEPHONE NUMBER *(include area code)*
Unclassified	Unclassified	Unclassified			

UNIFORMED SERVICES UNIVERSITY OF THE HEALTH SCIENCES
4301 JONES BRIDGE ROAD
BETHESDA, MARYLAND 20814-4712

DECEMBER 08, 2016

MEMORANDUM FOR STEVEN DURNING, M.D., PH.D., DEPARTMENT OF MEDICINE

SUBJECT: Uniformed Services University (USU) Institutional Review Board (IRB; FWA 0001628; DoD Assurance P60001) Approval of Protocol MED-83-3824 for Human Subjects Participation

Congratulations! The initial review for your No More Than Minimal Risk human subjects research protocol MED-83-3824, entitled "Developing Assessment Tools to Better Understand the Mechanisms of Clinical Reasoning in Military Medical Simulation," was reviewed and approved for execution on December 06, 2016 by Edmund Howe, M.D., J.D., Chair IRB #1 under the provision of 32 CFR 219.110(b)(1)Suppl.F(7). This approval will be reported to the USUHS IRB #1 scheduled to meet on January 12, 2016.

This project aims to explore, through video recorded simulated learning encounters, the nature of clinical reasoning and how features in a given situation may impact how physicians decide and act. The ultimate goal of this project is to develop an intervention to improve clinical reasoning performance and reduce medical errors. Up to 126 residents and physicians in primary care and surgical specialties are authorized to participate in this study.

Authorization to conduct protocol MED-83-3824 will automatically terminate on December 08, 2017. Annual review is required to maintain authorization to conduct this study. Please submit an application for continuing approval 60 days prior to your termination date.

You are required to submit amendments to this protocol, changes to the informed consent document (if applicable), adverse event reports, and other information pertinent to human research for this project. If you have questions regarding this action, or questions of a more general nature concerning human participation in research, please contact Micah Stretch at micah.stretch@usuhs.edu or (301) 295-9534.

Edmund G. Howe, M.D., J.D
Chair, IRB #1

Concur / Nonconcur

Arthur L. Kellerman, MD, MPH
Professor and Dean, School of Medicine

UNIFORMED SERVICES UNIVERSITY OF THE HEALTH SCIENCES
4301 JONES BRIDGE ROAD
BETHESDA, MARYLAND 20814
www.usuhs.mil

February 02, 2017

MEMORANDUM FOR DR STEVEN J DURNING P and R - Uniformed Services University of the Health Sciences (USUHS)

SUBJECT: USUHS IRB #1 (FWA 00001628; DoD Assurance P60001) Amendment Approval

Congratulations! The Modification for your No More Than Minimal Risk human subjects research protocol MED-83-3824, titled "Developing assessment tools to better understand the mechanisms of clinical reasoning in military medical simulation", was reviewed and approved for execution on February 02, 2017 by Edmund G. Howe, M.D., J.D. by Expedited Review under Category 7: Research on individual or group characteristics or behavior (including, but not limited to, research on perception, cognition, motivation, identity, language, communication, cultural beliefs or practices, and social behavior) or research employing survey, interview, oral history, focus group, program evaluation, human factors evaluation, or quality assurance methodologies . This approval will be reported to the USU IRB #1 scheduled to meet on February 09, 2017.

This modification adds Walter Reed National Military Medical Center as a performance site and Dr. Jeffrey LaRochelle as an Additional Investigator in the eIRB system. These modifications were previously approved in the original submission of the protocol to the IRB.

Authorization to conduct MED-83-3824 will automatically terminate on December 5, 2017. If you plan to continue data collection or analysis beyond this date, IRB approval for continuation is required. Please submit an application for continuing approval to the IRB Office 60 days prior to your termination date.

You are required to submit amendments to this protocol, changes to the informed consent document (if applicable), adverse event reports, and other information pertinent to human research for this project. No changes to this protocol may be implemented prior to IRB approval. If you have questions regarding this IRB action or questions of a more general nature concerning human subject participation in research, please contact Christopher Murphy at 301-319-0444 or christopher.murphy.ctr@usuhs.edu.

Edmund Howe, MD
Chair, USUHS IRB #1

UNIFORMED SERVICES UNIVERSITY OF THE HEALTH SCIENCES
4301 JONES BRIDGE ROAD
BETHESDA, MARYLAND 20814-4799
www.usuhs.mil

August 29, 2017

MEMORANDUM FOR DR STEVEN J DURNING, P&R- UNIFORMED SERVICES UNIVERSITY OF THE HEALTH SCIENCES

SUBJECT: USUHS IRB #1 (FWA 00001628; DoD Assurance P60001) Approval of Modification of Protocol MED-83-3824 for Human Subjects Participation

Congratulations! The Modification for your No More Than Minimal Risk human subjects research protocol MED-83-3824, entitled ***"Developing Assessment Tools to Better Understand the Mechanisms of Clinical Reasoning in Military Medical Simulation",*** was reviewed and approved for execution on August 29, 2017 by Dr. Edmund Howe MD, JD, Chair IRB #1, under the provision of 32 CFR 219.110(b)(1)Suppl. F(7). This approval will be reported to the USUHS IRB #1 scheduled to meet on September 14, 2017.

This project aims to explore the nature of clinical reasoning and how features in a given situation may impact how physicians decide and act. The objectives of this study include, a) use of existing, and novel measures to assess cognitive performance using two different genres of simulation-based learning environments; b) to examine theoretical linkages in a proposed conceptual model with potential moderators, both in vivo and in vitro; and c) to develop an intervention with the goal of enhancing our understanding of clinical reasoning and ultimately improving patient care.

Authorization to conduct protocol will automatically terminate on 12/05/2017. If you plan to continue data collection or analysis beyond this date, IRB approval for continuation is required.

You are required to submit amendments to this protocol, changes to the informed consent document (if applicable), adverse event reports, and other information pertinent to human research for this project. No changes to this protocol may be implemented prior to IRB approval. If you have questions regarding this IRB action or questions of a more general nature concerning human participation in research, please contact Chris Murphy at 301-319-0444 or christopher.murphy.ctr@usuhs.edu

Chris Murphy
IRB Analyst

UNIFORMED SERVICES UNIVERSITY OF THE HEALTH SCIENCES

4301 JONES BRIDGE ROAD
BETHESDA, MARYLAND 20814-4799
www.usuhs.mil

August 29, 2017

MEMORANDUM FOR DR STEVEN J DURNING, P&R- UNIFORMED SERVICES UNIVERSITY OF THE HEALTH SCIENCES

SUBJECT: USUHS IRB #1 (FWA 00001628; DoD Assurance P60001) Approval of Modification of Protocol MED-83-3824 for Human Subjects Participation

Congratulations! The Modification for your No More Than Minimal Risk human subjects research protocol MED-83-3824, entitled ***"Developing Assessment Tools to Better Understand the Mechanisms of Clinical Reasoning in Military Medical Simulation",*** was reviewed and approved for execution on August 29, 2017 by Dr. Edmund Howe MD, JD, Chair IRB #1, under the provision of 32 CFR 219.110(b)(1)Suppl. F(7). This approval will be reported to the USUHS IRB #1 scheduled to meet on September 14, 2017.

This project aims to explore the nature of clinical reasoning and how features in a given situation may impact how physicians decide and act. The objectives of this study include, a) use of existing, and novel measures to assess cognitive performance using two different genres of simulation-based learning environments; b) to examine theoretical linkages in a proposed conceptual model with potential moderators, both in vivo and in vitro; and c) to develop an intervention with the goal of enhancing our understanding of clinical reasoning and ultimately improving patient care.

Authorization to conduct protocol will automatically terminate on 12/05/2017. If you plan to continue data collection or analysis beyond this date, IRB approval for continuation is required.

You are required to submit amendments to this protocol, changes to the informed consent document (if applicable), adverse event reports, and other information pertinent to human research for this project. No changes to this protocol may be implemented prior to IRB approval. If you have questions regarding this IRB action or questions of a more general nature concerning human participation in research, please contact Chris Murphy at 301-319-0444 or christopher.murphy.ctr@usuhs.edu

Chris Murphy
IRB Analyst

UNIFORMED SERVICES UNIVERSITY OF THE HEALTH SCIENCES
4301 JONES BRIDGE ROAD
BETHESDA, MARYLAND
20814-4799
www.usuhs.edu

February 20, 2018

MEMORANDUM FOR DR STEVEN J DURNING, P&R- UNIFORMED SERVICES UNIVERSITY OF THE HEALTH SCIENCES

SUBJECT: USUHS IRB #1 (FWA 00001628; DoD Assurance P60001) Approval of Modification (Ref 894048) of Protocol MED-83-3824 for Human Subjects Participation

Congratulations! The Modification for your No More Than Minimal Risk human subjects research protocol MED-83-3824, entitled "Developing Assessment Tools to Better Understand the Mechanisms of Clinical Reasoning in Military Medical Simulation", was reviewed and approved for execution on February 20, 2018 by Dr. Edmund Howe MD, JD, Chair IRB #1, under the provision of 32 CFR 219.110(b)(1)Suppl. F(7). This approval will be reported to the USUHS IRB #1 scheduled to meet on March 8, 2018.

This project aims to explore the nature of clinical reasoning and how features in a given situation may impact how physicians decide and act. The objectives of this study include, a) use of existing, and novel measures to assess cognitive performance using two different genres of simulation-based learning environments; b) to examine theoretical linkages in a proposed conceptual model with potential moderators, both in vivo and in vitro; and c) to develop an intervention with the goal of enhancing our understanding of clinical reasoning and ultimately improving patient care.

This modification includes the following: This is for Modification #3 (Ref 894048) for the following: 1) Addition of working with The University of Texas Health Science Center (UTHSCA) at San Antonio for study Aim #2; 2) Addition of Fort Belvoir as a study site; 3) Addition of new three new study investigators Greg Condos, Luke Surry, and Abigail Konopask; 4) Correction of discrepancies in protocol documents; and 5) addition of The Henry M. Jackson Foundation for the Advancement of Military Medicine, who is funding the study, to the consent form.

Authorization to conduct protocol will automatically terminate on 12/05/2018. If you plan to continue data collection or analysis beyond this date, IRB approval for continuation is required.

You are required to submit amendments to this protocol, changes to the informed consent document (if applicable), adverse event reports, and other information pertinent to human research for this project. No changes to this protocol may be implemented prior to IRB approval. If you have questions regarding this IRB action or questions of a more general nature concerning human participation in research, please contact Maggie Pickerel at margaret.pickerel.ctr@usuhs.edu or 301-295-9813.

Margaret Pickerel

Margaret Pickerel
Senior IRB Coordinator

UNIFORMED SERVICES UNIVERSITY OF THE HEALTH SCIE C
4301 JONES BRIDGE ROAD
BETHESDA, MARYLAND
20814-4799
www.usuhs.ed
u

April 12, 2018

MEMORANDUM FOR STEVEN DURNING, M.D., PH.D., DEPARTMENT OF MEDICINE

SUBJECT: USU IRB (FWA 00001628; DoD Assurance P60001) Modification (ref# 902393) Approval for Protocol MED-83-3824 for Human Subjects Participation

Congratulations! The Modification (ref# 902393) for your No More Than Minimal Risk research protocol MED-83-3824, entitled "***Developing assessment tools to better understand the mechanisms of clinical reasoning in military medical simulation***," was reviewed and approved for execution on April 11, 2018 by Edmund G. Howe, M.D., J.D., Chair IRB pursuant to 32 CFR 219.110(b)(2). This approval will be reported to the USU IRB scheduled to meet on May 10, 2018.

The objectives of this study include a) use of existing, and novel measures to assess cognitive performance using simulation-based learning environments; b) to examine theoretical linkages in a proposed conceptual model with potential moderators; and c) to develop an intervention with the goal of enhancing our understanding of clinical reasoning and ultimately improving patient care.

The modification submission proposes the following changes to the protocol:

1. Addition of Brooke Army Medical Center (BAMC) as a new study site (Site #5).
2. Revision of consent form to reflect the addition of BAMC as a study site.
3. Moving Dr. Abigail Konopasky to section 3 of the eIRB protocol template as an Associate Investigator since she now has eIRB access.

Authorization to conduct protocol MED-83-3824 will automatically terminate on December 05, 2018. If you plan to continue data collection or analysis beyond this date, IRB approval for continuation is required. Please submit an application for continuing approval to the IRB Office 60 days prior to your termination date.

You are required to submit amendments to this protocol, changes to the informed consent document (if applicable), adverse event reports, and other information pertinent to human research for this project. No changes to this protocol may be implemented prior to IRB approval. If you have questions regarding this action or questions of a more general nature concerning human participation in research, Yaw Adomako-Ankomah, PhD at 301-295- 0428 or yaw.adomako-ankomah.ctr@usuhs.edu.

ADOMAKO-ANKOMAH.YAW.1541043442

Digitally signed by ADOMAKO-ANKOMAH.YAW.1541043442
Date: 2018.04.12 15:52:28

Yaw Adomako-Ankomah, PhD
IRB Analyst
USU Human Research Protection Program

UNIFORMED SERVICES UNIVERSITY OF THE HEALTH SCIE
- 041 31 J ON ESBRI OS1 AR
EODGONRATHAS, MAJ R
Y4L1- 2- 877
9 9 9 wu. suwme
.

May 08, 2018

MEMORANDUM FOR STEVEN DURNING, M.D., PH.D., DEPARTMENT OF MEDICINE

SUBJECT: USU Institutional Review Board (IRB) (FWA 00001628; DoD Assurance P60001) Approval of Amendment ref# 904072 to Protocol MED-83-3824 for Human Subjects Participation

Congratulations! The Amendment ref# 902393 for your No More Than Minimal Risk research protocol MED-83-3824, entitled "***Developing assessment tools to better understand the mechanisms of clinical reasoning in military medical simulation***," was reviewed and approved for execution on May 08, 2018 by Jeffrey L. Goodie, Ph.D., Vice Chair USU IRB under the provision of 32 CFR 219.110(b)(2). This approval will be reported to the USU IRB scheduled to meet on May 31, 2018.

The objectives of this study include a) use of existing, and novel measures to assess cognitive performance using simulation-based learning environments; b) to examine theoretical linkages in a proposed conceptual model with potential moderators; and c) to develop an intervention with the goal of enhancing our understanding of clinical reasoning and ultimately improving patient care.

The modification submission proposes the following changes to the protocol:

1. Addition of Megan Ohmer, Mark Haigney and Stephanie Ajuzie as Associate Investigators

Authorization to conduct protocol MED-83-3824 will automatically terminate on December 05, 2018. If you plan to continue data collection or analysis beyond this date, IRB approval for continuation is required. Please submit an application for continuing approval to the IRB Office 60 days prior to your termination date.

You are required to submit amendments to this protocol, changes to the informed consent document (if applicable), adverse event reports, and other information pertinent to human research for this project. No changes to this protocol may be implemented prior to IRB approval. If you have questions regarding this action or questions of a more general nature concerning human participation in research, Yaw Adomako-Ankomah, PhD at 301-295- 0428 or yaw.adomako-ankomah.ctr@usuhs.edu.

AR1 HAd1 2
AJ d1 HAGw
A
K wli - 14- 0- -
Y

Rqgal yn ugbhe Vm AR1 HAd1 2
AJ d1 HAGw AK wli - 14- 0- - Y
Rl an5Y41Lwi wL 105045YL 2
4- :44:

Yaw Adomako-Ankomah, PhD
IRB Analyst
USU Human Research Protection Program

UNIFORMED SERVICES UNIVERSITY OF THE HEALTH SCIENCES
4301 JONES BRIDGE ROAD
BETHESDA, MARYLAND
20814-4799
www.usuhs.edu

June 25, 2018

MEMORANDUM FOR STEVEN DURNING, M.D., PH.D., DEPARTMENT OF MEDICINE

SUBJECT: USU Institutional Review Board (IRB) (FWA 00001628; DoD Assurance P60001) Approval of Amendment ref# 905400 to Protocol MED-83-3824 for Human Subjects Participation

Congratulations! The Amendment ref# 905400 for your No More Than Minimal Risk research protocol MED-83-3824, entitled "***Developing assessment tools to better understand the mechanisms of clinical reasoning in military medical simulation***," was reviewed and approved for execution on June 21, 2018 by Edmund G. Howe, M.D., J.D., Chair IRB under the provision of 32 CFR 219.110(b)(2). This approval will be reported to the USU IRB scheduled to meet on July 12, 2018.

The objectives of this study include a) use of existing, and novel measures to assess cognitive performance using simulation-based learning environments; b) to examine theoretical linkages in a proposed conceptual model with potential moderators; and c) to develop an intervention with the goal of enhancing our understanding of clinical reasoning and ultimately improving patient care.

The Modification submission proposes the following changes to the approved protocol:

1. Corrective action for RE (ref# 904927); Revision of Section 12.4 (Special categories) of the eIRB Protocol Template to be consistent with the inclusion criteria language in Section 12.5, and in the IRB-approved 3204 protocol.

This action updates the approved eIRB Protocol Template to version 1.15.

Authorization to conduct protocol MED-83-3824 will automatically terminate on December 05, 2018. If you plan to continue data collection or analysis beyond this date, IRB approval for continuation is required. Please submit an application for continuing approval to the IRB Office 60 days prior to your termination date.

You are required to submit amendments to this protocol, changes to the informed consent document (if applicable), adverse event reports, and other information pertinent to human research for this project. No changes to this protocol may be implemented prior to IRB approval. If you have questions regarding this action or questions of a more general nature concerning human participation in research, Yaw Adomako-Ankomah, PhD at 301-295- 0428 or yaw.adomako-ankomah.ctr@usuhs.edu.

ADOMAKO-ANKOMAH.YAW.1541043442

Digitally signed by ADOMAKO-ANKOMAH.YAW.1541043442 Date: 2018.06.25 10:09:44 -04'00'

Yaw Adomako-Ankomah, PhD IRB Analyst

Learning to Care for Those in Harm's Way

v.4/27/18

UNIFORMED SERVICES UNIVERSITY OF THE HEALTH SCIENCES
4301 JONES BRIDGE ROAD
BETHESDA, MARYLAND
20814-4799
www.usuhs.edu

August 10, 2018

MEMORANDUM FOR STEVEN DURNING, M.D., PH.D., DEPARTMENT OF MEDICINE

SUBJECT: USU Institutional Review Board (IRB) (FWA 00001628; DoD Assurance P60001) Approval of Amendment ref# 905608 to Protocol MED-83-3824 for Human Subjects Participation

Congratulations! The Amendment ref# 905608 for your No More Than Minimal Risk research protocol MED-83-3824, entitled "***Developing assessment tools to better understand the mechanisms of clinical reasoning in military medical simulation***," was reviewed and approved for execution on August, 2018 by Edmund G. Howe, M.D., J.D., Chair IRB under the provision of 32 CFR 219.110(b)(2). This approval will be reported to the USU IRB scheduled to meet on August 23, 2018.

The objectives of this study include a) use of existing, and novel measures to assess cognitive performance using simulation-based learning environments; b) to examine theoretical linkages in a proposed conceptual model with potential moderators; and c) to develop an intervention with the goal of enhancing our understanding of clinical reasoning and ultimately improving patient care.

The Modification submission proposes the following changes to the approved protocol:

1. Conversion of the protocol to the new multi-site format in eIRB.
2. Removal of the following personnel from the Core Protocol:
    a. Luke Surry
    b. Jeffrey Mikita

This action updates the approved eIRB Protocol Template to version 1.16.

Authorization to conduct protocol MED-83-3824 will automatically terminate on December 05, 2018. If you plan to continue data collection or analysis beyond this date, IRB approval for continuation is required. Please submit an application for continuing approval to the IRB Office 60 days prior to your termination date.

You are required to submit amendments to this protocol, changes to the informed consent document (if applicable), adverse event reports, and other information pertinent to human research for this project. No changes to this protocol may be implemented prior to IRB approval. If you have questions regarding this action or questions of a more general nature concerning human participation in research, Yaw Adomako-Ankomah, PhD at 301-295- 0428 or yaw.adomako-ankomah.ctr@usuhs.edu.

ADOMAKO-ANKOMAH.YAW.1541043442

Digitally signed by ADOMAKO-ANKOMAH.YAW.1541043442
Date: 2018.08.10 16:16:55 -04'00'

Yaw Adomako-Ankomah, PhD
IRB Analyst
USU Human Research Protection Program

UNIFORMED SERVICES UNIVERSITY OF THE HEALTH SCIENCES
4301 JONES BRIDGE ROAD
BETHESDA, MARYLAND
20814-4799
www.usuhs.edu

September 23, 2019

MEMORANDUM FOR STEVEN DURNING, M.D., PH.D., DEPARTMENT OF MEDICINE

SUBJECT: USU Institutional Review Board (IRB) (FWA 00001628; DoD Assurance P60001) Approval of Amendment ref# 917404 to Protocol MED-83-3824 for Human Subjects Participation

The Amendment ref# 917404 for your No More Than Minimal Risk research protocol MED-83-3824, entitled "***Developing assessment tools to better understand the mechanisms of clinical reasoning in military medical simulation***," was reviewed and approved for execution on September 18, 2019 by Edmund G. Howe, M.D., J.D., Chair IRB under the provision of 32 CFR 219.110(b)(2). This approval will be reported to the USU IRB scheduled to meet on September 26, 2019.

The objectives of this study include a) use of existing, and novel measures to assess cognitive performance using simulation-based learning environments; b) to examine theoretical linkages in a proposed conceptual model with potential moderators; and c) to develop an intervention with the goal of enhancing our understanding of clinical reasoning and ultimately improving patient care.

The MODIFICATION submission proposes the following changes to the approved protocol:

1. Removal of Dr. Gregory Condos from the study
2. Addition of Jerusalem Merkebu as Associate Investigator

The following study documents were reviewed:

1. EIRB Modification Form - (Version 12.2)
2. EIRB Protocol Template - (Version 1.20)
3. Personnel support documents (CITI, COI, CV) for Jerusalem Merkebu

Authorization to conduct protocol MED-83-3824 will automatically terminate on December 05, 2019. If you plan to continue data collection or analysis beyond this date, IRB approval for continuation is required. Please submit an application for continuing approval to the IRB Office 60 days prior to your termination date.

You are required to submit amendments to this protocol, changes to the informed consent document (if applicable), adverse event reports, and other information pertinent to human research for this project. No changes to this protocol may be implemented prior to IRB approval. If you have questions regarding this action or questions of a more general nature concerning human participation in research, Yaw Adomako-Ankomah, PhD at 301-295- 0428 or yaw.adomako-ankomah.ctr@usuhs.edu.

ADOMAKO-ANKOMAH.YAn .1b41043442 Digitally signed by ADOMAKO-ANKOMAH.YAn .1b41043442 Date: 2019.09.23 11:06:24 -04'00'

Yaw Adomako-Ankomah, PhD
IRB Analyst
USU Human Research Protection Program

UNIFORMED SERVICES UNIVERSITY OF THE HEALTH SCIENCES
4301 JONES BRIDGE ROAD
BETHESDA, MARYLAND
20814-4799
www.usuhs.edu

December 20, 2019

MEMORANDUM FOR STEVEN DURNING, M.D., PH.D., DEPARTMENT OF MEDICINE

SUBJECT: USU Institutional Review Board (IRB) (FWA 00001628; DoD Assurance P60001) Approval of Amendment ref# 921808 to Protocol MED-83-3824 for Human Subjects Participation

The Amendment ref# 921808 for your Minimal Risk research protocol MED-83-3824, entitled "***Developing assessment tools to better understand the mechanisms of clinical reasoning in military medical simulation***," was reviewed and approved for execution on December 20, 2019 by Edmund G. Howe, M.D., J.D., Chair IRB under the provision of 32 CFR 219.110(b)(2). This approval will be reported to the USU IRB scheduled to meet on January 09, 2019.

The objectives of this study include a) use of existing, and novel measures to assess cognitive performance using simulation-based learning environments; b) to examine theoretical linkages in a proposed conceptual model with potential moderators; and c) to develop an intervention with the goal of enhancing our understanding of clinical reasoning and ultimately improving patient care.

This action approves the following modifications to the protocol:

1. Addition of Dr. Michael Soh to the protocol as an Associate Investigator.

The following study documents were reviewed:

1. EIRB Modification Form - (Version 13.0)
2. EIRB Protocol Template - (Version 1.22)
3. Personnel support documents - Soh

Authorization to conduct protocol MED-83-3824 will automatically terminate on December 05, 2020. If you plan to continue data collection or analysis beyond this date, IRB approval for continuation is required. Please submit an application for continuing approval to the IRB Office 60 days prior to your termination date.

You are required to submit amendments to this protocol, changes to the informed consent document (if applicable), adverse event reports, and other information pertinent to human research for this project. No changes to this protocol may be implemented prior to IRB approval. If you have questions regarding this action or questions of a more general nature concerning human participation in research, Yaw Adomako-Ankomah, PhD at 301-295- 0428 or yaw.adomako-ankomah.ctr@usuhs.edu.

ADOMAKO-ANKOMAH.YAW.1541043442

Digitally signed by ADOMAKO-ANKOMAH.YAW.1541043442 Date: 2019.12.20 15:47:01 -05'00'

Yaw Adomako-Ankomah,
PhD IRB Analyst
USU Human Research Protection Program

UNIFORMED SERVICES UNIVERSITY OF THE HEALTH SCIENCES
4301 JONES BRIDGE ROAD
BETHESDA, MARYLAND
20814-4799
www.usuhs.edu

November 20, 2018

MEMORANDUM FOR STEVEN DURNING, M.D., PH.D., DEPARTMENT OF MEDICINE

SUBJECT: USU Institutional Review Board (IRB) (FWA 00001628; DoD Assurance P60001) Approval of Protocol MED-83-3824 for Human Subjects Participation

Congratulations! The Continuing Review ref# 907726 for your No More Than Minimal Risk human subjects research protocol MED-83-3824, entitled " ***Developing assessment tools to better understand the mechanisms of clinical reasoning in military medical simulation***," was reviewed and approved for execution on November 19, by Edmund G. Howe, M.D., J.D., Chair IRB under the provision of 32 CFR 219.110(b)(1)Suppl. F(7). This approval will be reported to the USU IRB scheduled to meet on November 29, 2018.

The objectives of this study include a) use of existing, and novel measures to assess cognitive performance using simulation-based learning environments; b) to examine theoretical linkages in a proposed conceptual model with potential moderators; and c) to develop an intervention with the goal of enhancing our understanding of clinical reasoning and ultimately improving patient care.

Authorization to conduct protocol MED-83-3824 will automatically terminate on December 05, 2019. If you plan to continue data collection or analysis beyond this date, IRB approval for continuation is required. Please submit an application for continuing approval to the IRB Office 60 days prior to your termination date.

You are required to submit amendments to this protocol, changes to the informed consent document (if applicable), adverse event reports, and other information pertinent to human research for this project. No changes to this protocol may be implemented prior to IRB approval. If you have questions regarding this IRB action or questions of a more general nature concerning human participation in research, please contact Yaw Adomako-Ankomah, PhD at 301-295- 0428 or yaw.adomako-ankomah.ctr@usuhs.edu.

ADOMAKO-ANKOMAH.YAW.1541043442

Digitally signed by ADOMAKO-ANKOMAH.YAW.1541043442
Date: 2018.11.21 09:54:17 -05'00'

Yaw Adomako-Ankomah, PhD
IRB Analyst
USU Human Research Protection Program

UNIFORMED SERVICES UNIVERSITY OF THE HEALTH SCIENCES
4301 JONES BRIDGE ROAD
BETHESDA, MARYLAND
20814-4799
www.usuhs.edu

December 04, 2019

MEMORANDUM FOR STEVEN DURNING, M.D., PH.D., DEPARTMENT OF MEDICINE

SUBJECT: USU Institutional Review Board (IRB) (FWA 00001628; DoD Assurance P60001) Approval of Protocol MED-83-3824

The Continuing Review ref# 920152 for your No More Than Minimal Risk human subjects research protocol MED-83-3824, entitled " ***Developing assessment tools to better understand the mechanisms of clinical reasoning in military medical simulation***," was reviewed and approved for execution on December 03, 2019 by Edmund G. Howe, M.D., J.D., Chair IRB under the provision of 32CFR 219.110(b)(1)Suppl. F(7). This approval will be reported to the USU IRB scheduled to meet on December 12, 2019.

The objectives of this study include a) use of existing, and novel measures to assess cognitive performance using simulation-based learning environments; b) to examine theoretical linkages in a proposed conceptual model with potential moderators; and c) to develop an intervention with the goal of enhancing our understanding of clinical reasoning and ultimately improving patient care.

Authorization to conduct protocol MED-83-3824 will automatically terminate on December 05, 2020. If you plan to continue data collection or analysis beyond this date, IRB approval for continuation is required. Please submit an application for continuing approval to the IRB Office 60 days prior to your termination date.

You are required to submit amendments to this protocol, changes to the informed consent document (if applicable), adverse event reports, and other information pertinent to human research for this project. No changes to this protocol may be implemented prior to IRB approval. If you have questions regarding this IRB action or questions of a more general nature concerning human participation in research, please contact Yaw Adomako-Ankomah, PhD at 301-295- 0428 or yaw.adomako-ankomah.ctr@usuhs.edu.

ADOMAK
ANKOMAH.
W.1541043442

Digitally signed by ADOMAKO-ANKOMAH.YAW.1541043442
Date: 2019.12.04 13:42:19 -05'00'

Yaw Adomako-Ankomah,
PhD IRB Analyst
USU Human Research Protection Program

UNIFORMED SERVICES UNIVERSITY OF THE HEALTH SCIE
4301 JONES BRIDGE ROAD
BETHESDA, MARYLAND
20814-4799
www.usuhs.ed
u

May 14, 2018

MEMORANDUM FOR STEVEN DURNING, M.D., PH.D., DEPARTMENT OF MEDICINE

SUBJECT: USU Institutional Review Board (IRB) (FWA 00001628; DoD Assurance P60001) Approval of Amendment ref# 904273 to Protocol MED-83-3824 for Human Subjects Participation

Congratulations! The Amendment ref# 904273 for your No More Than Minimal Risk research protocol MED-83-3824, entitled "***Developing assessment tools to better understand the mechanisms of clinical reasoning in military medical simulation***," was reviewed and approved for execution on May 14, 2018 by Edmund G. Howe, M.D., J.D., Chair IRB under the provision of 32 CFR 219.110(b)(2). This approval will be reported to the USU IRB scheduled to meet on May 31, 2018.

The objectives of this study include a) use of existing, and novel measures to assess cognitive performance using simulation-based learning environments; b) to examine theoretical linkages in a proposed conceptual model with potential moderators; and c) to develop an intervention with the goal of enhancing our understanding of clinical reasoning and ultimately improving patient care.

The modification submission proposes the following changes to the protocol:

1. Removal of Fort Belvoir Community Hospital as a study site.

This action approves the following study documents: 1) EIRB Protocol Template – (Version 1.14); and 2) Revised consent 2.5 (English) (version 2.11).

Authorization to conduct protocol MED-83-3824 will automatically terminate on December 05, 2018. If you plan to continue data collection or analysis beyond this date, IRB approval for continuation is required. Please submit an application for continuing approval to the IRB Office 60 days prior to your termination date.

You are required to submit amendments to this protocol, changes to the informed consent document (if applicable), adverse event reports, and other information pertinent to human research for this project. No changes to this protocol may be implemented prior to IRB approval. If you have questions regarding this action or questions of a more general nature concerning human participation in research, Yaw Adomako-Ankomah, PhD at 301-295- 0428 or yaw.adomako-ankomah.ctr@usuhs.edu.

ADOMAKO-ANKOMAH.YA
W.1541043442

Digitally signed by ADOMAKO-ANKOMAH.YAW.1541043442
Date: 2018.05.14 13:26:12

Yaw Adomako-Ankomah, PhD
IRB Analyst
USU Human Research Protection Program

UNIFORMED SERVICES UNIVERSITY OF THE HEALTH SCIENCES
4301 JONES BRIDGE ROAD
BETHESDA, MARYLAND
20814-4799
www.usuhs.edu

November 21, 2017

MEMORANDUM FOR STEVEN J. DURNING, M.D., PH.D, DEPARTMENT OF MEDICINE

SUBJECT: USU IRB #1 (FWA 00001628; DoD Assurance P60001) Continuing Review of Protocol MED-83-3824

Congratulations! The Continuing Review for your no more than minimal risk human subjects research protocol MED-83-3824, entitled "Developing Assessment Tools to Better Understand the Mechanisms of Clinical Reasoning in Military Medical Simulation" was reviewed and approved for continuation on November 20, 2017 by Edmund G. Howe, M.D., J.D., Chair IRB #1 under the provision of 32 CFR 219.110(b)(1)Suppl.F(7). This approval will be reported to the USU IRB #1 scheduled to meet on December 14, 2017.

This project aims to explore, through video recorded simulated learning encounters, the nature of clinical reasoning and how features in a given situation may impact how physicians decided and act. The ultimate goal of this project is to develop an intervention to improve clinical reasoning performance and reduce medical errors. Up to 126 residents and physicians in primary care and surgical specialties are authorized to participate in this study.

Authorization to conduct protocol MED-83-3824 will automatically terminate on December 05, 2018. If you plan to continue data collection or analysis beyond this date, IRB approval for continuation is required. Please submit an application for continuing approval to the IRB Office 60 days prior to your termination date.

You are required to submit amendments to this protocol, changes to the informed consent document (if applicable), adverse event reports, and other information pertinent to human research for this project. No changes to this protocol may be implemented prior to IRB approval. If you have questions regarding this IRB action or questions of a more general nature concerning human participation in research, please contact Micah Stretch at 301-295-0819 or micah.stretch@usuhs.edu.

STRETCH.MICAH.R.10829108

Digitally signed by STRETCH.MICAH.R.1082910863
DN: c=US, o=U.S. Government, ou=DoD, ou=PKI, ou=OSD, cn=STRETCH.MICAH.R.1082910863
Reason: I am the author of this document Date: 2017.11.21 09:11:45 -05'00'

Micah R. Stretch, M.A., J.D.
Senior IRB Coordinator
Exemption Determination Official

The Use of Think-Aloud Reflections to Examine Learners Experiences in Live and Video-Based Simulation Contexts: A Comparison Study

Abigail Konopasky[1], Steven J. Durning[1], Divya Ramani[1], Megan Ohmer[1], Anthony Artino[1], Alexis Battista[1]

[1]Department of Medicine, Graduate Programs in Health Professions Education

Uniformed Services University, Bethesda, MD

INTRODUCTION

Purpose

To use "think-aloud" reflections to explore differences in learners' experiences between video and live scenario simulation contexts.

Background

- Recent meta-analyses argue that to advance the field of simulation, we must seek to better understand what works, **for whom** and **under what circumstances**.[1-2]
- Studies suggest that learner experiences differ across **simulated learning methods** (e.g., live scenario, video case) and **designs** (e.g., problem-solving, procedural skill): however, these studies primarily focus on performance outcomes rather than learner **experiences**.[3-4]
- Clarifying learner experiences may help deepen our understanding of what works and for whom by revealing **learning processes**.
- **Think-aloud reflection** offers us a potential window into understanding learning processes.[5]

Hypothesis

Simulation context differences → learner reflection differences

Research Questions

1. Does reflection differ in video and live simulation contexts? If so, how?
2. What differences does reflection reveal in learner experiences of each context?

METHODS

Design

Prospective Mixed-Methods Quasi Experiment[6]

Procedure

24 attending and resident physicians were quasi-randomly assigned to either:

1. watch two **video cases** *or* participate in a **live scenario simulation** of two cases (diabetes mellitus and coronary artery disease for both video and live).
2. Participants were then asked to "**think aloud**[3]" about their reasoning towards their diagnosis as they either:
   - Re-watched the **video case** *or*
   - Watched a video of their own **live simulation** performance

Figure 1. Procedure Workflow

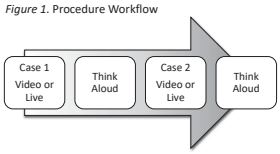


Table 1. *Participant Details*

	Video	Live
Specialty		
Internal Medicine	*N* = 11	*N* = 6
Family Medicine	*N* = 1	*N* = 1
Surgery	*N* = 1	*N* = 4
Other Demographics	Mean (SD)	Mean (SD)
Age	38 (9.9)	37 (11.6)
Years in Practice	10 (9.6)	9 (12.4)

Data Coding

1. Transcribed think-aloud reflections were hand coded for **reconsiderations**: moments when practitioners questioned their own or the video doctor's choices or thought processes.
2. Think-aloud reflections were also automatically coded for **linguistic markers** (Table 2) by the Linguistic Analysis and Word Count (LIWC) program.[7]

Table 2. *Linguistic Markers Coded*

LIWC Category	Example	Use
First-person pronouns	*I, we*	Attentional focus on self
Cognitive processing	*consider, explain*	Thinking styles
Affect	*worry, great*	Emotionality

Data Analysis

- Video and live scenario simulation conditions were **compared** using chi-square or *t*-tests as appropriate.

RESULTS: RECONSIDERATION

- Reconsiderations (i.e., moments when participants questioned their own or the video doctor's choices) were **more likely** to occur in the **live scenario** condition than in the video condition. For example:

"When I was doing, like, review of systems, I would've tried to pinpoint more, like, thyroid problems or things that, like more differential questions. I honed in on a diagnosis, and I focused a lot of my questions toward this diagnosis." (Live scenario, diabetes case)

- While **95% of live scenarios** had reconsiderations like this, only **50% of video-based scenarios** did.
- Also, only live scenario participants reconsidered **management** decisions like this:

[Regarding patient activity at home] *"I probably should have said more explicitly like, 'Don't exert yourself until we can get this more worked up.'"* (Live scenario, coronary artery disease case)

Table 3. *Reconsideration*

	Video	Live	X^2
No reconsideration codes	13	1	11.9* (Φ = .5)
Some reconsideration codes	13	21	
Total	22	26	

Note: *$p < .001$

RESULTS: LINGUISTIC MARKERS

- In the **live scenario** condition, participants used more **"I" pronouns** and words indicating **cognitive processes** (e.g., *think*) than in the video condition. For example:

"At this point I am trying to tease out whether or not this is something that is specifically related to exercise, if starting and stopping starts and stops the pain, or if it is something that happens to occur at the same time. But all of his answers pushed towards it very much linked to the exercise." (Live scenario, coronary artery disease case, LIWC coded words underlined)

- Emotional markers (e.g., *worry, great*), however, were not significantly different across conditions.

Table 4. *Linguistic Makers as a Percentage of Total Words Uttered*

	Video *m* (SD)	Live *m* (SD)	t
First-person pronouns	2.7%[i] (2)	5.5%[i] (2.3)	3.2** (d = 1.3)
Cognitive processing	16%[i] (4.1)	19.9%[i] (4)	2.4* (d = 1)
Affect	4.1%[i] (1.4)	3.6%[i] (.9)	1 (d = .4)

Note: [i]LIWC results are given as percentage of total words uttered
*$p < .05$, ** $p < .01$ 50

DISCUSSION

Participants in the **live scenario** simulations **reconsidered** their choices and actions **more than** participants assigned to video-based cases**;** also reconsidering **management** choices and actions.

- This adds insight into **what kinds** of learning occurs in these different simulated learning methods which, in turn, can be used to inform instructional or curricular design choices.[1-2]

Participants in the **live scenarios** focused more on the self (*I/me*) and cognitive processes (e.g., weighting diagnostic differences with words like *but* or *if* and connecting thoughts with words like *related*).

- The findings for reconsideration and linguistic marker analyses, taken together, suggest that thinking aloud stimulates **metacognitive processes** (e.g., thinking about one's thinking). Additionally, these data also suggest that through the use of think-aloud methodology, participants demonstrate a range of reflective processes even **without a faculty or peer guide**.

The use of linguistic markers and reconsideration coding represents a **direct measure** of participant reflections, offering a new way to examine reflection and feedback in healthcare simulation.

- Future research could apply these measures to examine the impact of differing approaches to guided debriefing, such as advocacy inquiry or debriefing for meaningful learning.[8-9]

PRACTICE APPLICATION IDEA

Consider using a **think-aloud reflection** following a live or video-based simulation **in place of a facilitated debriefing**. For example, if participants are scheduled to engage in two scenarios during a session, schedule one of those scenarios to use a think-aloud protocol to support reflection.

Think-Aloud Implementation Guidelines[10]:

1. To promote a safe environment, allow participants to view and complete their think aloud in **privacy**.
2. **Rehearse** with the designated facilitator(s) (ideally someone who doesn't supervise the individual). It can be very hard to not speak while participants think aloud.
3. Give students/learners an opportunity to **warm up**.
4. Plan for **enough time**. We recommend five minutes for setting up the video and warming up plus the total scenario time as a minimum.

REFERENCES

1. Cook, D. A., Hamstra, S. J., Brydges, R., Zendejas, B., Szostek, J. H., Wang, A. T., ... & Hatala, R. (2013). Comparative effectiveness of instructional design features in simulation-based education: systematic review and meta-analysis. *Medical teacher*, *35*(1), e867-e898.
2. Cook, D. A., Hatala, R., Brydges, R., Zendejas, B., Szostek, J. H., Wang, A. T., ... & Hamstra, S. J. (2011). Technology-enhanced simulation for health professions education: a systematic review and meta-analysis. *Jama*, *306*(9), 978-988.
3. Bong, C. L., Lightdale, J. R., Fredette, M. E., & Weinstock, P. (2010). Effects of simulation versus traditional tutorial-based training on physiologic stress levels among clinicians: a pilot study. *Simulation in Healthcare*, *5*(5), 272-278.
4. Bearman, M. (2003). Is virtual the same as real? Medical students' experiences of a virtual patient. *Academic Medicine*, *78*(5), 538-545.
5. Ericsson KA, Simon HA. How to study thinking in everyday life: Contrasting think-aloud protocols with descriptions and explanations of thinking. *Mind, Culture, and Activity*. 1998;*5*(3):178-86.
6. Ericsson, K. A., & Simon, H. A. (1998). How to study thinking in everyday life: Contrasting think-aloud protocols with descriptions and explanations of thinking. *Mind, Culture, and Activity*, *5*(3), 178-186.
7. Tausczik, Y. and Pennebaker, J. (2009). The Psychological Meaning of Words: LIWC and Computerized Text Analysis Methods. *Journal of Language and Social Psychology*, 29(1), pp.24-54.
8. Rudolph, J. W., Simon, R., Dufresne, R. L., & Raemer, D. B. (2006). There's no such thing as "nonjudgmental" debriefing: a theory and method for debriefing with good judgment. *Simulation in Healthcare*, *1*(1), 49-55.
9. Dreifuerst, K. T. (2011). Debriefing for meaningful learning©: a reflective strategy to foster clinical reasoning. *Clinical Simulation in Nursing*, *7*(6), e250.
10. Battista, A., Konopasky, A., Ramani, D., Ohmer, M., Mikita, J., Howle, A., Krajnik, S., Torre, D. and Durning, S. (2018). Clinical Reasoning in the Primary Care Setting: Two Scenario-Based Simulations for Residents and Attendings. *MedEdPORTAL*, 14.

Examining the Influence of Simulation Context on Learners' Post-Simulation Reflections

Abigail Konopasky[1], Steven J. Durning[1], Divya Ramani[1], Megan Ohmer[1], Anthony R. Artino Jr[1], Alexis Battista[1]

[1]Uniformed Services University of the Health Sciences, Bethesda, Maryland, USA

Introduction

Background

- Best-practice guidelines indicate that post-simulation reflection is an integral component simulation-based learning (SBL) (Decker et al., 2013).
- Research also suggests that learner experiences differ across simulation typologies (e.g., live scenario, video cases) and designs (e.g., problem solving, procedural skill) (e.g., Bong et al., 2010).

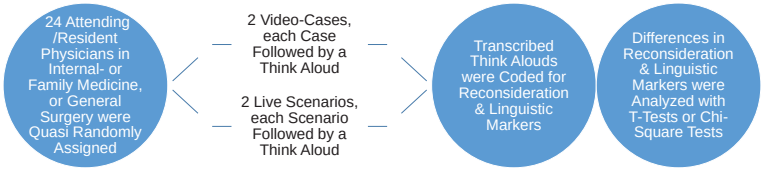
Purpose

- This study uses think aloud" reflections (Ericsson & Simon, 1998) to compare the impact of pre-recorded video cases and live scenario-based simulations on learners' reflections.

Research Questions

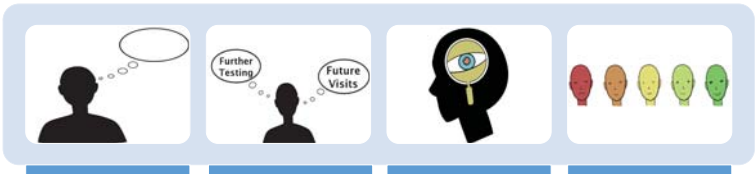
1. Does reflection as a learning tool differ across contexts? If so, how?
2. What differences do think-aloud reflections reveal in learner experiences across contexts?

Methods



Results

Simulation Typology Can Influence *What* Learners Reflect On



95% of Live-Scenario Participants Reconsidered Their Choices Only 50% of Video Participants Reconsidered Their Own or the Video Doctors Choices	Only Live-Scenario Participants Reconsidered Management Decisions	Live Scenario Participants Used More "I" Pronouns and Cognitive Processing Words	Emotional Markers Were Not Significantly Different Across Conditions

Note: *Reconsiderations* are defined as moments when participants questioned their own or the video doctor's choices. The use of *first-person pronouns (e.g.,* "I") indicates whether attention is on the self. *Cognitive processing* words (e.g., consider, explain), indicate thinking styles (Tausczik & Pennebaker, 2009). *Typology* is defined as "the classification of different educational methods or equipment" (Lopreiato, 2016 p. 40). Characteristics of *management reasoning* were derived from Cook, Sherbino, & Durning (2018).

Conclusions

Summary

- There were more and different kinds of reflections for live-scenario participants, with a focus on the self's thoughts and choices, suggesting that differences in simulation context can influence learners' experiences and post-simulation reflection.

Take Home Message

- As one of the first inferential test of varying simulation contexts, this study reveals important differences in how simulation context may influence learners' reflection content.

Interested in Learning More? Contact Us!

Abigail Konopasky, PhD abigail.konopasky.ctr@usuhs.edu @AKonopasky
Alexis Battista, PhD alexis.battista.ctr@usuhs.edu @abbattist

USU Uniformed Services University

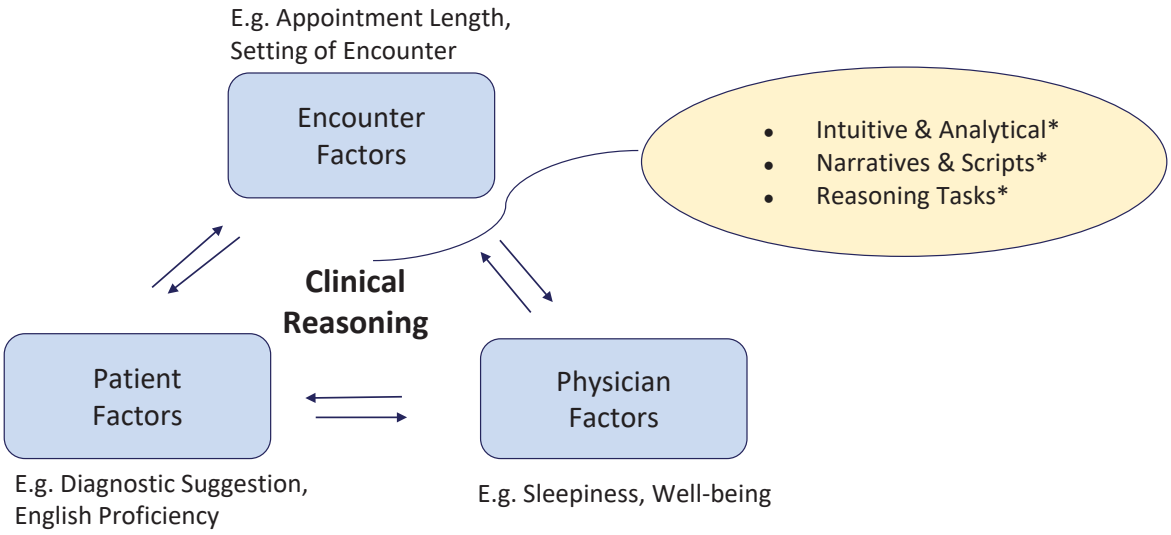
Emergent Clinical Reasoning During Think Alouds: How Physicians Reflect on their Own and Others' Practices in Live and Video Simulation

Abigail Konopasky, PhD; Alexis Battista, PhD; Megan Ohmer, Divya Ramani, MS; Anthony R. Artino, Jr., PhD; Steven Durning, MD, PhD

Uniformed Services University

This work was funded by a Congressionally Directed Medical Research Program, Joint Program Committee-1 Grant.

The Complexity of Clinical Reasoning



*Custers, 2015; Juma & Goldszmidt, 2017; Mattingly, 1991; McBee et al., 2016; Norman, 2009

USU
Uniformed Services University

Representation and Authenticity

*

- What features and levels of authenticity support clinical reasoning? (Issenberg, Ringsted, Ostergaard, & Diekmann, 2011)
- Focus on *functional* alignment and emotional engagement rather than physical realism (DeMaria, et al., 2010; Hamstra, Brydges, Hatal, Zendejas, & Cook, 2014)

*Image from Henrik Hagtvedt's work at: http://hagtvedt.com/history.html

Thinking Aloud

- Think-aloud reflection: speaking thoughts aloud without description or explanation (Ericsson & Simon, 1984, 1993)
- This *type* of reflection appears to be a reasonable measure of thinking (Durning et al., 2013; Ericsson & Simon, 1998; Fox, Ericsson, & Best, 2011)
- Effective method in prior studies of assessing clinical reasoning (Burbach, Barnason, & Thompson, 2015; Durning et al., 2012a,b; Funkesson, Anbacken, & Ek, 2007)

METHODS

Participant Details

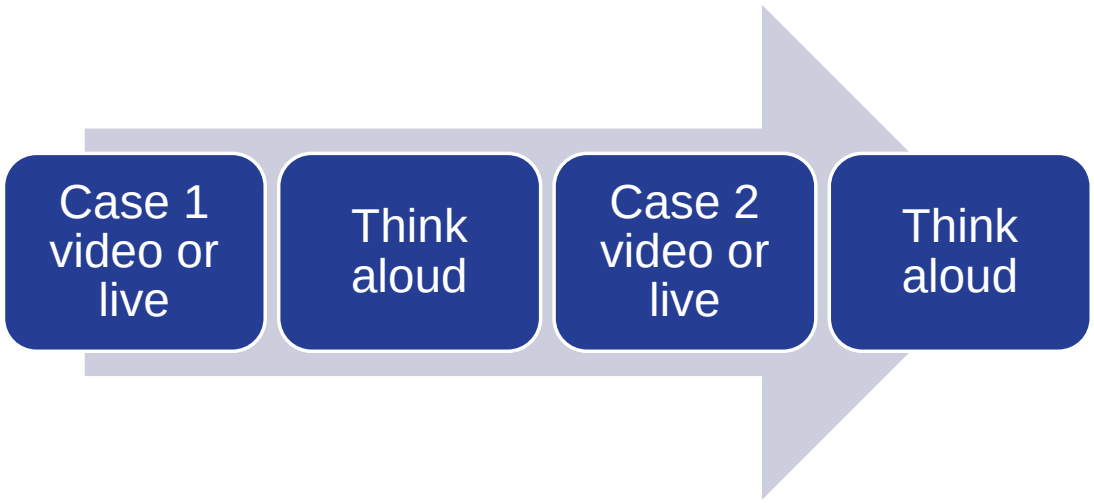
Total *N* = 14

	Video	Live
Internal Medicine	*N* = 4	*N* = 4
Family Medicine	*N* = 1	*N* = 1
Surgery	*N* = 1	*N* = 3

	Video		Live	
	Mean (SD)	Range	Mean (SD)	Range
Age (6 female; 8 male)	34 years (9.2)	28-52 years	33 years (5.3)	28-43 years
Years in practice	7 years (9.8)	1-26 years	3.5 years (4.1)	1-13 years

Study Design

Quasi-random assignment into video or live condition and then:



USU
Uniformed Services University

Reconsideration Analysis

- Think alouds are a form of reflection.
- A central dimension of reflection is its *iterative* nature: revisiting beliefs or experiences to produce new understandings (Mann, Gordon, & MacLeod, 2009).
- Our interest: these new understandings, i.e., when practitioners questioned prior understanding of beliefs, choices or thought processes → instances of ***reconsideration***

Research Questions

1. What types of thoughts do physicians reconsider during think-aloud reflections?
2. How do reconsiderations differ, if at all, between video and live conditions?
3. How often do physicians reconsider thoughts in the video compared to live condition?

RESULTS

RQ1: Types of Reconsiderations

Three categories: diagnostic, practice improvement, prior case comparison

Diagnostic	Practice Improvement	Prior Case Comparison
I should've asked about sweating—if he gets sweaty. (Sc)	I also feel like sometimes I leave out important things, especially in my review of systems. (Sc)	I think with this interaction, maybe because it was my second, but also with chest pain, I felt overall a little bit more confidence. (Sc)
I definitely want to do a fundoscopic exam on her though. (V)	I would have gotten an interpreter…because I think even though it doesn't change what the diagnosis is, it would have improved the rapport. (V)	I think, like the last patient, I'd be concerned with his compliance. (V)

USU
Uniformed Services University

RQ2: Descriptive Comparison

How do the structure and content of reconsiderations differ in live vs. video?

	Live	Video
Action	I was going to check on her nose and her ears.	So detailed cardiovascular exam would be important for this patient: blood pressure and heart rate and looking for the evidence of thrush in the mouth.
Reason	but she didn't have any complaints, so that was lower on my list.	[n/a]

RQ2: Descriptive Comparison

How do the structure and content of reconsiderations differ in live vs. video?

	Live	Video
Action	When I was going through review of systems, I would've tried to pinpoint more, like, thyroid problems or things like that. Like more differential questions.	I'd like him to have asked her what her problems were even if it has been a while since she had them.
Reason	I honed in on a diagnosis and focused a lot of my questions toward this diagnosis.	[n/a]

RQ3: Quantitative Comparison

How *often* do participants reconsider actions in live versus video?

- Varying lengths of think-alouds, so we examined *presence* versus *absence* of reconsiderations
  - We found a statistically significant difference btwn live and video

	Condition		χ^2	*p*
	Live	Video		
No Reconsiderations	3	15	9.63	.002
Some Reconsiderations	21	13		

*Note that a larger sample (*N* = 52 cases, 26 participants) was used for this analysis

USU
Uniformed Services University

DISCUSSION

Discussion

- Participants **reconsider** multiple aspects of the complex task of clinical reasoning across video and live conditions
  - Might be a type of **cognitive forcing** strategy: analytically reconsidering result of unconscious pattern recognition
  - Think-alouds: good reflective tool for the **complexity** of clinical reasoning
- Reconsiderations in live scenarios tend to be more **frequent** and more **complex**, involving both action and reasons behind it
  - Neither live nor video necessarily *better*, but different reflective foci
- Comparisons between first and second case (diabetes and angina): new way of looking at how **context** impacts performance: participants compare *across* content areas

USU
Uniformed Services University

Limitations

- Small sample drawn from a single site
- Single source of data
- Retrospective (vs. concurrent) think alouds

Conclusions

- Authenticity matters
- Reconsiderations = potentially useful markers of clinical reasoning for instruction and assessment
- Prior patients may impact clinical reasoning

Thank You!

Abigail Konopasky, PhD (abigail.konopasky.ctr@usuhs.edu)

Alexis Battista, PhD (alexis.battista.ctr@usuhs.edu)

Megan Ohmer (megan.ohmer@usuhs.edu)

Divya Ramani, MS (divya.ramani.ctr@usuhs.edu)

Anthony R. Artino, Jr., PhD (anthony.artino@usuhs.edu)

Steven J. Durning, MD, PhD (steven.durning@usuhs.edu)

For more info on think-alouds and development of the scenarios:
Battista A, Konopasky A, Ramani D, Ohmer M, Mikita J, Howle A, Krajnik S, Torre D, Durning S. Clinical Reasoning in the Primary Care Setting: Two Scenario-Based Simulations for Residents and Attendings. *MedEdPORTAL.* In Press

USU
Uniformed Services University

Disclosure Slide

- *Abigail Konopasky*
  - None
- *Alexis Battista*
  - None
- *Divya Ramani*
  - None
- *Megan Ohmer*
  - None
- *Anthony Artino*
  - None
- *Steven Durning*
  - ABIM Consultant
  - Academic Medicine Deputy Editor

The Effect of Contextual Factors on Clinical Reasoning: A Mixed Methods Study Examining Outcome and Process

Abigail Konopasky, PhD; Alexis Battista, PhD; Divya Ramani, MS; Megan Ohmer; Anthony Artino, PhD; Steven J Durning, MD, PhD

USU
Uniformed Services University

This work was funded by a Congressionally Directed Medical Research Programs' Joint Program Committee-1 Grant.

Objectives

- To understand **context specificity**
- To understand how contextual factors impact **performance**
- To explore outcomes and processes of **clinical reasoning**

Image from http://inmyownterms.com/finding-the-right-context-for-a-term/

USU
Uniformed Services University

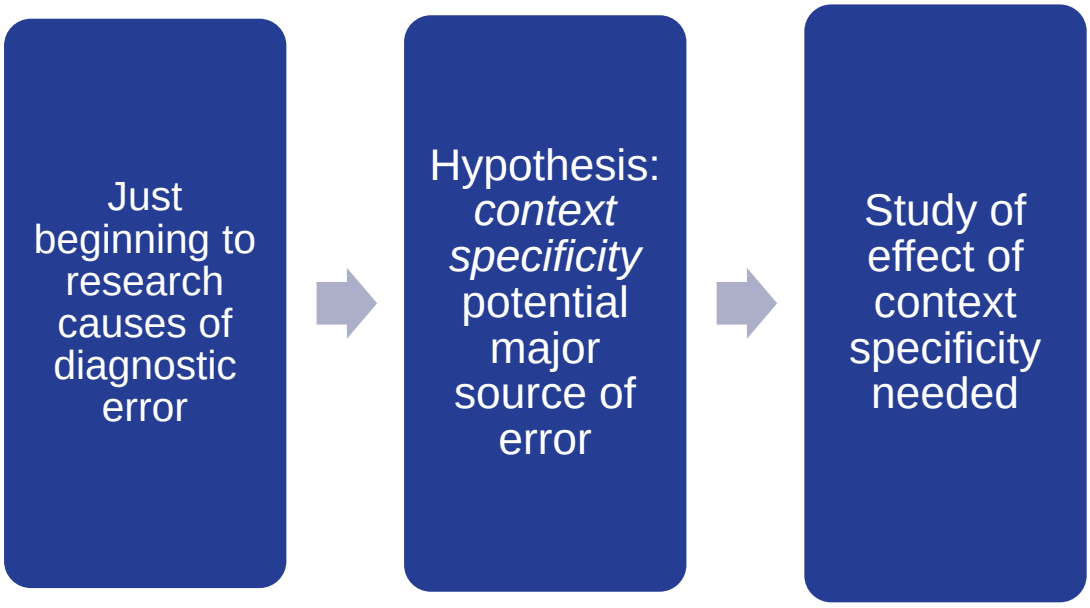
BACKGROUND

USU
Uniformed Services University

Diagnostic Error

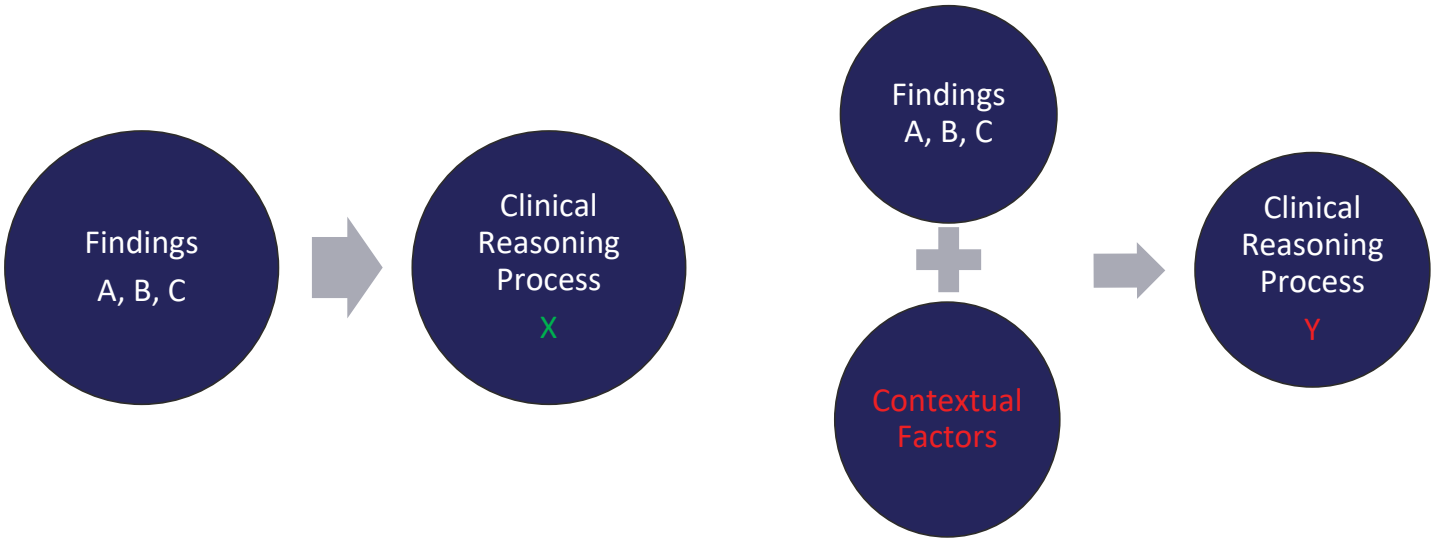
- “Most people will experience at least one diagnostic error in their lifetime, sometimes with devastating consequences.”
- Diagnostic error is a leading cause of death in US

Addressing Diagnostic Error



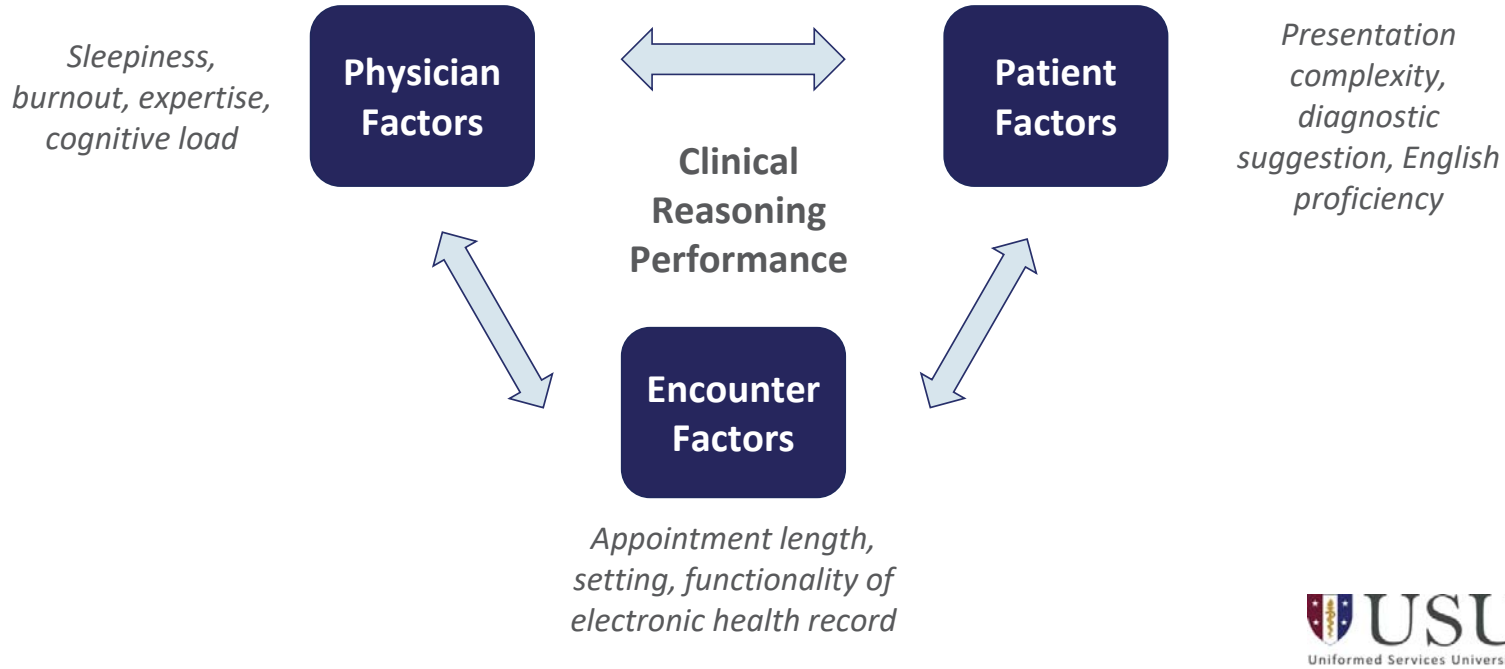
USU
Uniformed Services University

Context Specificity



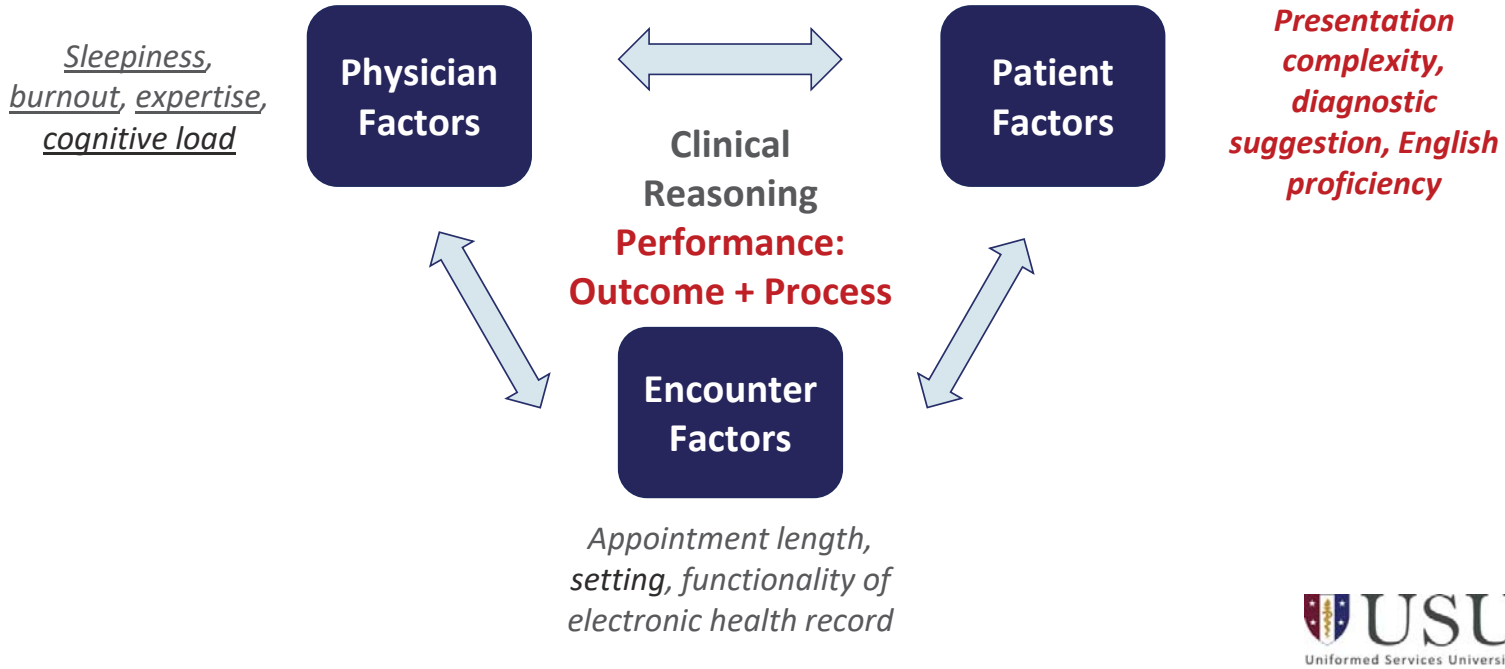
USU
Uniformed Services University

Situated Cognition Theory



USU
Uniformed Services University

Contextual Factors & Clinical Reasoning



USU
Uniformed Services University

METHODS

USU
Uniformed Services University

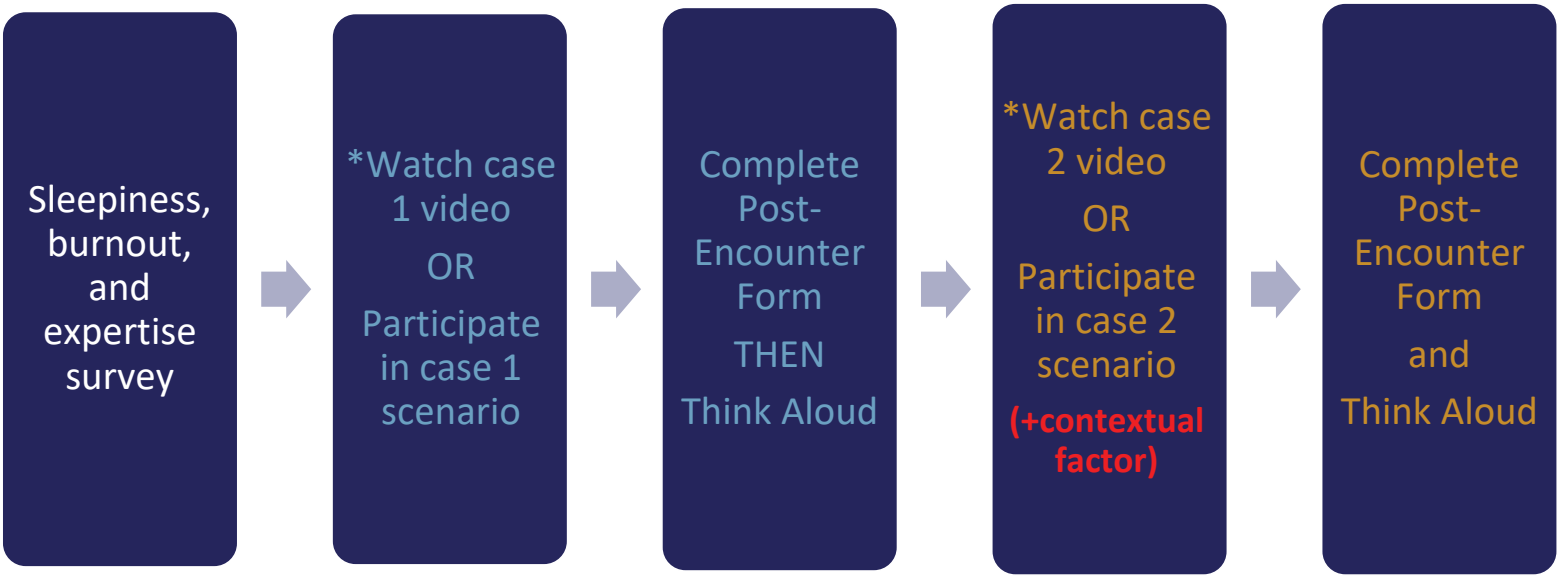
Participant Details

29 participants in internal medicine, family medicine, and surgery

	Mean (SD)	Range
Age (8 female; 21 male)	36 years (11)	25-66 years
Years in practice	9.2 years (11.4)	1-39 years

USU
Uniformed Services University

Procedure



*Order of these is varied: half of participants have video/scenario with no contextual factors first and half receive it second

USU
Uniformed Services University

Hypotheses & Research Questions

Quantitative Hypotheses

- Outcome scores (DPEF) will be **lower** in the presence of contextual factors.
- Cognitive load ratings will be **higher** in the presence of contextual factors

Qualitative Research Questions

- Are there distinct patterns of language use (i.e., contextual factor mentions, hedgers, subject pronouns) in the think-aloud transcripts across conditions?

RESULTS

Qualitative Results

1. Mention of contextual factors

- “But yeah, I feel like this part goes on too long. But it was just because he was so excited that he fixed his medical problem!” (+CF)

2. Emotional language and doubt

- “So this is really bothering her and he seems pretty callous about it” (+CF)

3. Comparison to participant’s practice

- [The doctor’s late and the patient’s upset] ”I’ve been there” (+CF)

Qualitative Results

4. **Hedging (qualifying statements): Contextual factor vs. diagnostic process**
   - "He brought up acid reflux. And it *seemed like* he was very excited that he had solved his problem and that this wasn't something scary, and I was *trying to like*… validate this? Because, you know, it *totally possibly could be*, but at this point I was very concerned that it was cardiac."(+CF)
   - So *I think* the first thing, is *kinda*, he's talking about pain in the center of his chest." (-CF)
5. **Generic *you/we*: Not enough information**
   - "*You* just don't get the history from her, though, that this was going on that long, so it's not necessarily consistent with the rest of her story, so it makes me wonder." (+CF)

Quantitative Results

	No Contextual Factors M (SD)	Contextual Factors M (SD)	*t*-test (significance)
Additional interview questions (0-10)	5 (2.6)	4.6 (3.2)	.8 (*d* = .14)
Additional exam items (0-10)	4.3 (2.2)	2.8 (2.1)	3.3** (*d* = .7)
Differential diagnosis (0-6)	4.2 (1.3)	4 (1.2)	.4 (*d* = .16)
Problem list (0-2)	1.5 (.3)	1.3 (.3)	2* (*d* = .67)
Leading diagnosis (0-2)	1.8 (.4)	1.4 (.4)	4.9*** (*d* = 1)
Supporting evidence (0-10)	9.4 (2.3)	6.8 (4)	2.7* (*d* = .8)
Cognitive load (1-10)	5.7 (1.5)	6.3 (1.5)	1.8 (*d* = .4)

*$p < .05$, **$p < .01$, ***$p < .001$

USU
Uniformed Services University

CONCLUSION

Conclusions

- Contextual factors significantly affect **performance**. We see this through:
  - Language markers (mention of CF, hedgers, generic *you*) in presence of CF
  - Impaired diagnostic performance in presence of CF
- Novel use of **theoretical** model (situated cognition) and **measures** (e.g., linguistics) to track effect of contextual factors
- Important step towards reducing **diagnostic error** through a better understanding of context specificity

Thank you!

- Feel free to contact me with thoughts and questions!
  - abigail.konopasky.ctr@usuhs.edu

BACKUP SLIDES

Study Design

PARTICIPANTS WERE ASSIGNED TO ONE OF THREE GROUPS:

Group A	Group B	Group C
Video	*Video*	*Live Scenario*
-Angina (-CF)	-Diabetes mellitus (-CF)	Diabetes mellitus (-CF)
-Diabetes mellitus (limited English)	-Angina (presentation complexity)	Angina (diagnostic suggestion)
N = 11	*N = 7*	*N = 11*

Data Sources

Quantitative

Diagnostic Post-Encounter Form (DPEF)

- Open-ended 6-item form asking about diagnosis & treatment

Cognitive Load Score

- Single-item rating of cognitive load on scale of 1-10

Qualitative

Think-Aloud Interview

- Rewatch video of case or rewatch self in scenario
- Asked to “think aloud” as to how reached diagnosis
- Minimal to no cuing from researcher

USU
Uniformed Services University

JPG1 Study Overview

Phase 1
- Development of videos and scenarios with and without contextual factors

Phase 2
- Collection of data from ~80 participants either watching or participating in cases

Phase 3
- Design, carry out, and assess intervention for effect of contextual factors on reasoning

Case Specificity in Clinical Reasoning: A Qualitative Case Study of Conditional Reasoning Processes

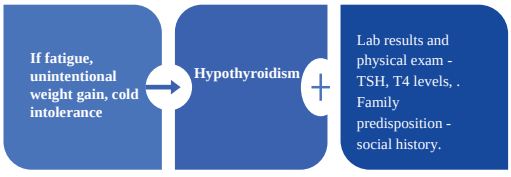
Divya Ramani MS, Abigail Konopasky PhD, Alexis Battista PhD, Anthony Artino PhD, & Steven Durning, MD,PhD
Health Professions Education Division, Department of Medicine, Uniformed Services University of the Health Sciences, Bethesda, MD

ABSTRACT

As part of a larger study of clinical reasoning, we examined how five physicians from different specialities (one family medicine attending, two internal medicine attendings, one surgery resident, and one internal medicine resident) reasoned through two video cases (medical condition A and medical condition B) when asked to "think aloud" about their reasoning. We analyzed 10 think-alouds (two per physician) and coded them for instances of *conditional reasoning*, where conditional reasoning refers to logical reasoning based on if/then statements. The findings suggest differences in reasoning processes as well as a need for further investigation of conditional reasoning processes.

METHODS

- Participants were asked to 'think aloud' about their reasoning processes while watching each of two videos (medical cases A and B).
- These sessions resulted in two 5-7 minute audio recordings for each participant (10 think-alouds total), which were then transcribed verbatim.
- We coded the think-aloud transcripts for participants' *conditional reasoning* processes: identifying and connecting antecedents (symptoms) and consequents (potential diagnoses), and using evidence (e.g., personal and family history, labs) to support or refute the connection between them (Sternberg,2006).



RESULTS

This qualitative case study involved five physicians (see Table 1). Think-aloud analyses revealed differences in physicians' conditional reasoning processes for the two medical cases, for instance, arriving at the diagnosis early in the process, versus towards the end, versus not being able to decide on an exact diagnosis at all.

We also found evidence of case specificity, which refers to variability of clinical reasoning performance dependent on case content: participants noted the diagnosis for condition A almost immediately, while there was variability in when diagnosis was noted for condition B, despite a similar level of difficulty.

Table 1: Participant details

	Gender	Speciality	Date of Graduation
A	Male	Family Medicine	1992
B	Female	Internal Medicine	2009
C	Male	Internal Medicine	2017
D	Male	Surgery	2017
E	Male	Internal Medicine	1994

DISCUSSION

- Examining clinical reasoning can help reduce medical errors and improve diagnosis (Norman & Eva, 2010), particularly across content areas, where *case specificity* (variability of clinical reasoning performance across cases) can lead to differences in clinical reasoning (ten Cate & Durning, 2018).
- As part of a larger, mixed-methods, clinical reasoning study, this qualitative analysis of five think alouds revealed the importance of attending to the individual components of clinical arguments. These components varied for different types of cases and for different individuals, even when those individuals had similar levels of training.
- These findings suggest that to improve diagnostic and management performance, instructors and researchers should further describe and characterize these conditional reasoning differences as a step towards understanding clinical decision making and preventing medical errors related to diagnosis.

CONCLUSION

Need for further investigation of conditional reasoning processes, which may be a helpful tool for understanding case specificity and possibly other aspects of physicians' reasoning.

REFERENCES

Norman, G. R., Eva, K., W.(2010). Diagnostic error and clinical reasoning. *Medical Education, 44*, 94-100.doi:10.1111/j.1365-2923.2009.03507.x
ten Cate, O., & Durning, S. J. (2018). Understanding Clinical Reasoning from Multiple Perspectives: A Conceptual and Theoretical Overview. In *Principles and Practice of Case-based Clinical Reasoning Education* (pp. 35-46). Springer, Cham.
Norman, G. (2005). Research in clinical reasoning: past history and current trends. *Medical education, 39*, 418-427. doi:10.1111/j.1365-2929.2005.02127.x
Stenberg,R.J.(2006).Cognitive Psychology,CA,*Wadsworth Cengage learning*

ACKNOWLEDGEMENTS

This qualitative study is one part of a larger study entitled: "Developing Assessment Tools to Better Understand the Mechanisms of Clinical Reasoning in Military Medical Simulation," funded by...
JPC-1, Congressionally Directed Medical Research Programs,
Award #: NH83382416

Title: Uncovering patterns of uncertainty across clinical reasoning tasks

Authors: Ramani, D., Soh, M., Merkebu, J., Mcbee, E., Ratcliffe, T., Konopasky, A., Artino, A., Durning, S.
Uniformed Services University Of Health Science, Bethesda, Maryland

Research Questions (24)

What is the relative frequency of markers of uncertainty across framing, diagnostic, management, and reflection tasks as clinicians think aloud through their reasoning process?

Background and Relevance (109)

Limitations in clinicians' clinical reasoning--encompassing framing, diagnostic, management, and reflection tasks[3]--have been linked to diagnostic error.[1] One challenge associated with clinical reasoning is *uncertainty*. While every clinical encounter inherently entails uncertainty, examining *when* it arises may help us better understand its role in clinical reasoning. Although research has begun to explore uncertainty's effect on clinical reasoning outcomes,[2,3] we understand little about the role of uncertainty across specific clinical reasoning *tasks*.[4] In order to eventually mitigate the detrimental effects of uncertainty and reduce diagnostic error, we must first examine precisely *where* it emerges in the clinical reasoning process. That is the purpose of this exploratory study.

Design and Methods (58)

This qualitative study examined how 20 practicing physicians reasoned as they "thought aloud" towards a diagnosis and management plan while watching a videotaped clinical encounter. The think alouds were coded for a) markers of uncertainty (e.g. "maybe," "probably," "could"),[5,6,7] and b) clinical reasoning tasks.[8] We looked for overlap between these markers and tasks.

Results (33)

Analysis revealed 220 uncertainty markers overlapping with clinical reasoning tasks. Of these 220 overlaps, approximately 29% (63) fell within framing, 60% (131) in diagnosis, 9% (20) in management, and 3% (6) in reflection.

Conclusion (49)

These findings suggest that physicians may treat framing, diagnosis, management, and reflection tasks differently. While uncertainty may function as a "cognitive space" for physicians to reconsider and double-check their judgments, future research should examine uncertainty's role in these four task categories and whether they hinder or facilitate patient care.

Total word count: 273/300

References:

1. ten Cate O, Durning SJ. Understanding Clinical Reasoning from Multiple Perspectives: A Conceptual and Theoretical Overview. 2018:35-46. doi:10.1007/978-3-319-64828-6_3
2. Zaat JOM, van Eijk JTM. General practitioners' uncertainty, risk preference, and use of laboratory tests. *Med Care*. 1992:846-854.
3. O'Riordan M, Aktürk Z, Ortiz JMB, et al. Dealing with uncertainty in general practice: an

essential skill for the general practitioner. *Qual Prim Care*. 2011;19(3).
4. Goldszmidt M, Minda JP, Bordage G. Developing a unified list of physicians' reasoning tasks during clinical encounters. Acad Med. 2013;88(3):390-394.
5. Konopasky AW, Ramani D, Ohmer M, et al. It Totally Possibly Could Be: How A Group Of Military Physicians Reflect On Their Clinical Reasoning In The Presence Of Contextual Factors. Mil Med.
6. Levinson P, Brown P, Levinson SC, Levinson SC. Politeness: Some Universals in Language Usage. Vol 4. Cambridge university press; 1987.
7. Bhise V, Rajan SS, Sittig DF, Morgan RO, Chaudhary P, Singh H. Defining and measuring diagnostic uncertainty in medicine: a systematic review. J Gen Intern Med. 2018:1-13.
8. Juma S, Goldszmidt M. What physicians reason about during admission case review. Adv Heal Sci Educ. 2017;22(3):691-711.

An Introductory Workshop for Activity and Linguistic Analysis of Video in Healthcare Simulation

As you arrive we ask that you use the sticky notes at your table to answer the following questions for your context & goal setting.

1. How are you currently using your video-based data (e.g., video-supported debriefing, quality assurance, program evaluation, research, etc.)?

2. What would you like to learn to do during today's workshop (e.g., learn some easy ways to analyze video, learn how to chunk video for analysis, etc.)?

USU
Uniformed Services University

An Introductory Workshop for Activity and Linguistic Analysis of Video in Healthcare Simulation

Alexis Battista, PhD & Abigail Konopasky, PhD[2]

USU
Uniformed Services University

Objectives

Goal: *Practical* tools for using video in teaching, assessment, faculty development, research, and more.

1. Identify practical ways participants can use their video-recordings to enhance teaching, assessment, faculty development or research.

2. Practice identifying common activity markers of learning or performance in video-recorded simulation data using structured coding tools.

3. Practice identifying common linguistic markers of performance in video-recorded simulation data using structured coding tools.

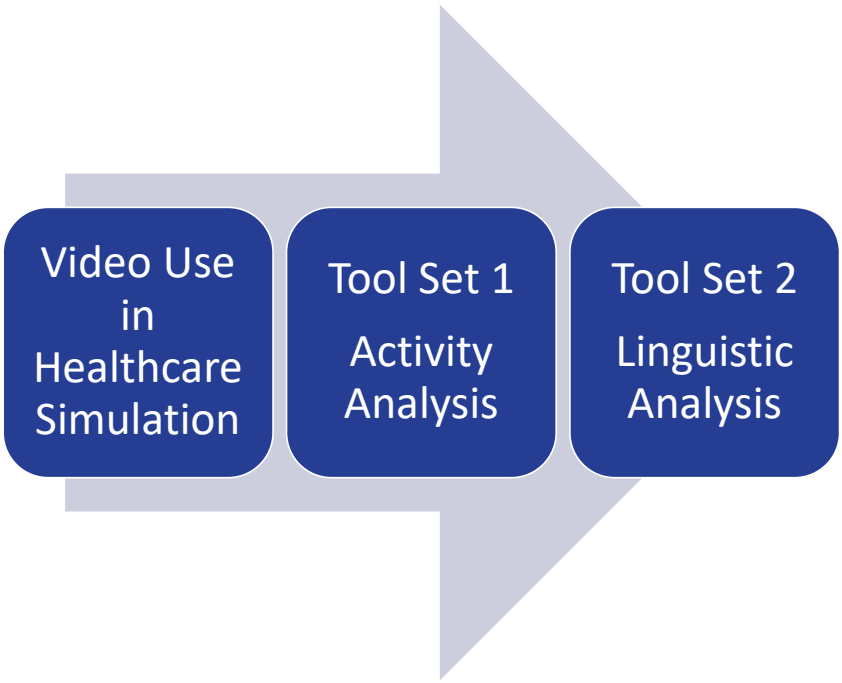
Disclosures

The presenters were supported by a grant from the JPC - 1, CDMRP - Congressionally Directed Medical Research Program (# NH83382416).

Disclaimer

The views expressed in this workshop are those of the authors and do not necessarily reflect the official position or policy of the US Government, Department of Defense, Department of the Navy, or the Uniformed Services University.

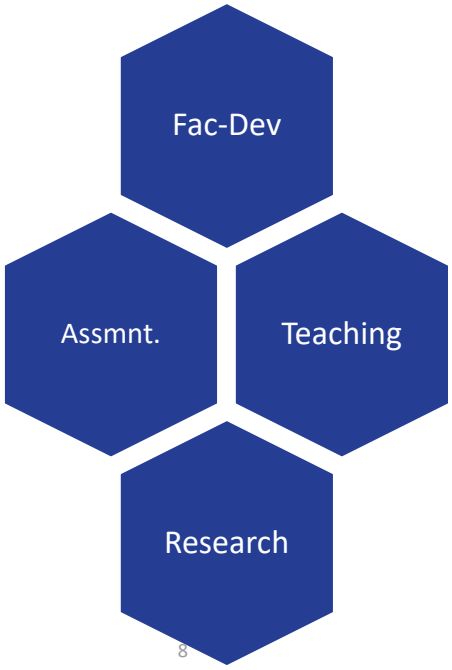
Session Workflow



USU
Uniformed Services University

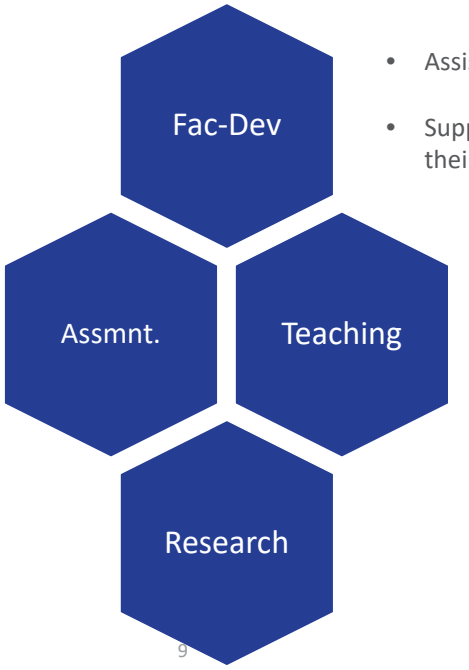
VIDEO USE IN HEALTHCARE SIMULATION

Some Common/Practical Uses of Video



USU
Uniformed Services University

Some Common/Practical Uses of Video

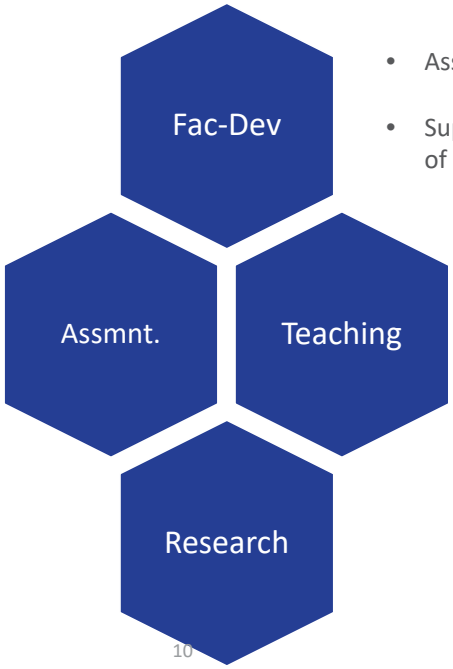


- Assist in evaluation of faculty
- Support faculty self-evaluation of their own teaching practices

USU
Uniformed Services University

Some Common/Practical Uses of Video

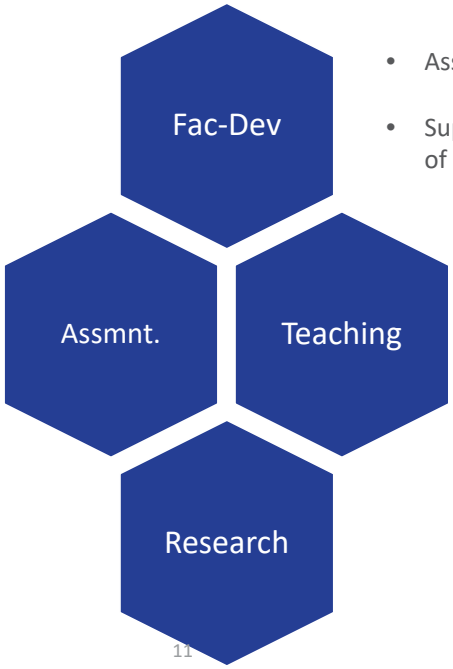
- Assist in evaluation of faculty
- Support faculty self-evaluation of their own performances



- Support asynchronous assessment or blinded assessment
- Serve as a back-up for revisiting assessment related sessions

USU
Uniformed Services University

Some Common/Practical Uses of Video



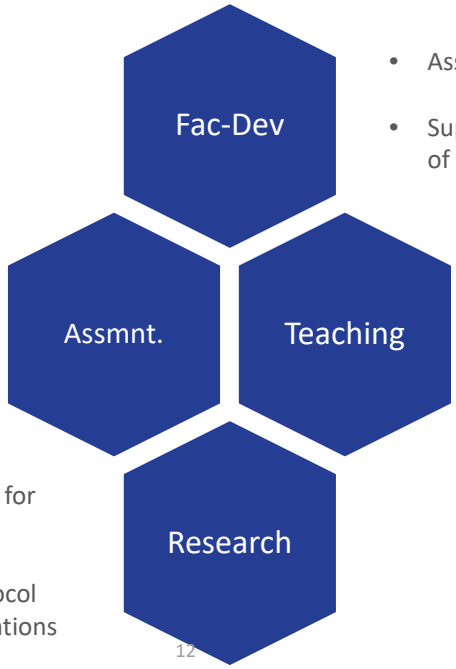
- Assist in evaluation of faculty
- Support faculty self-evaluation of their own performances

- Support asynchronous assessment or blinded assessment
- Serve as a back-up for revisiting assessment related sessions

- Video-assisted debriefing/reflection
- Provide modeling or guidance to students (especially for procedural skills training)

USU
Uniformed Services University

Some Common/Practical Uses of Video



- Assist in evaluation of faculty
- Support faculty self-evaluation of their own performances

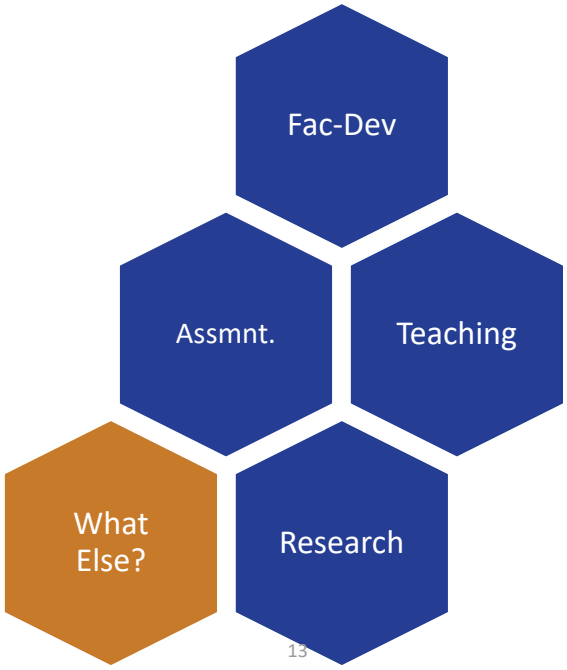
- Support asynchronous assessment or blinded assessment
- Serves as a back-up for revisiting assessment related sessions

- Video-Assisted Debriefing/Reflection
- Modeling or Provide Student Guidance

- Primary source of data for analysis
- Support research protocol implementation evaluations

USU
Uniformed Services University

In Sum: Video Data Use is Diverse



USU
Uniformed Services University

However...

Video data can also be overwhelming to analyze.

TOOL SET #1
ACTIVITY ANALYSIS

What is (isn't) Activity Analysis

Getting **beyond** checklists or scoring performance. Instead…

- Systematic approach to examining complex learning environments/situations.
- Helps account for the precise sequence of events that can help you understand patterns of behavior across a class/group/cohort.
- Also helps reveal gaps in what wasn't done.
- Can include very detailed video annotation/coding or be done by hand focusing on a very defined/focused set of information.
- Usually done most successfully *after* the simulated encounters are completed.

Getting Started with Activity Analysis

Step One:

1. Get a sense of what was recorded.

2. Check video recording quality
   - Check sound quality
   - Are the important people all in frame?

3. Record Basic Video Demographics
   - Who is in the video
   - What clinical roles are represented
   - How long is the video
   - Who is reviewing

Step 1 – What was Recorded?

Inpatient Rapid Response Scenario

- Insert Video Clip Here

IMPRESSIONS/ THOUGHTS?

Step 2: Choosing Who/What to Focus On

- Insert Still Image Here

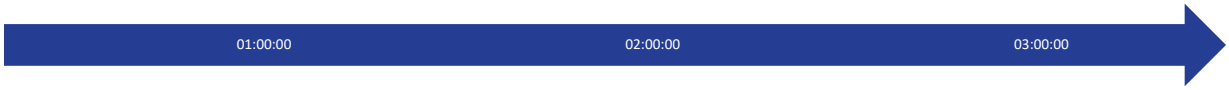
Simulation-based encounters are inherently complex so you have to be picky.

Coding ‘What and How’ Using an Existing Coding Scheme

Code	Operational Definition
Structured Interventions	Activities that participants perform that are governed by a set of predetermined rules.
Tools/Artifacts Used	Physical items that are present in the simulated setting that form the system that subjects may interact with or utilize to achieve their goals.
Social Interactions	Interactions participants have with others in the simulated context, including peers, faculty, and standard participants (e.g., patient, patient’s support person, anesthesiologist).

Recording Activities Using a Timeline

*Accounting for *time* and *sequence* while making it easier to compare more than one performance to another.



USU
Uniformed Services University

Step 3: Coding Structured Interventions

“Activities that participants perform that are governed by a set of predetermined rules.”

USU
Uniformed Services University

Your Turn to Code Structured Interventions

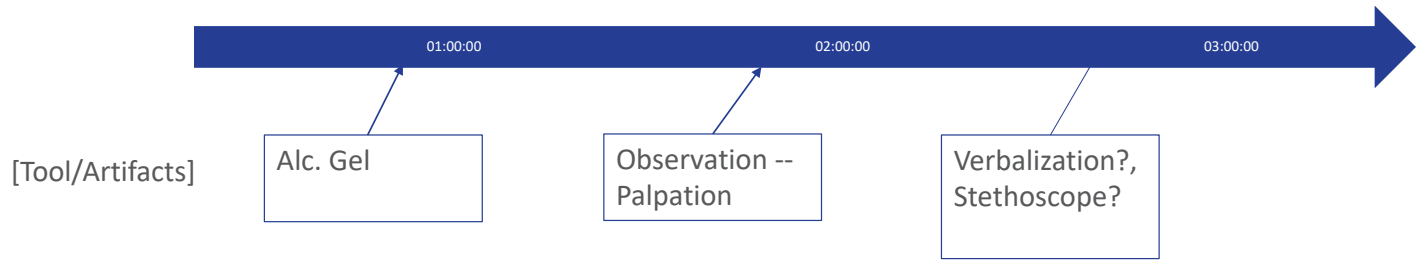
“Activities that participants perform that are governed by a set of predetermined rules.”

- Insert Video Clip Here

IMPRESSIONS/ THOUGHTS?

Step 4: Coding for Tool/Artifact Use

“Physical items subjects may interact with or utilize to achieve their goals.”



USU
Uniformed Services University

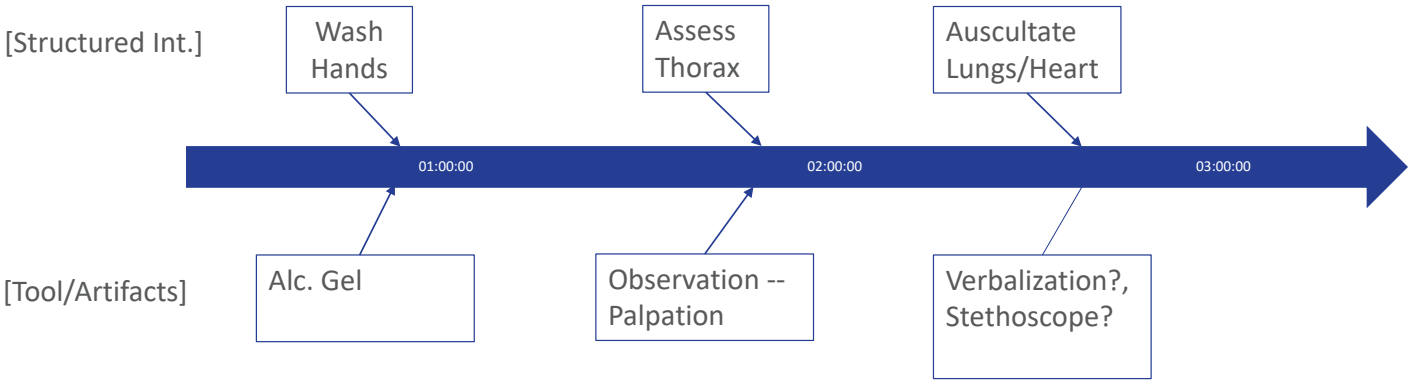
Your Turn to Code for Tool/Artifact Use

“Physical items subjects may interact with or utilize to achieve their goals.”

- Insert Video Clip Here

IMPRESSIONS/ THOUGHTS?

Pulling it Together



USU
Uniformed Services University

Summary of Tool Kit #1

1. Take the time to get to know what's been recorded.
2. Work with an existing coding scheme to help you focus.
3. Pick a focus.
4. Be prepared to make multiple passes to help you stay focused.
5. Chunk your video analysis into smaller blocks of time to limit fatigue or feeling overwhelmed.

TOOL SET #2
LINGUISTIC ANALYSIS

What is(n't) Linguistic Analysis?

AND HOW COULD IT POSSIBLY BE HELPFUL TO ME??

Linguistic analysis is NOT:

- Correct grammar (and annoyingly correcting others' grammar!)
- Latin and Greek roots
- Translation into foreign languages
- Examining accents or dialects

What is(n't) Linguistic Analysis?

AND HOW COULD IT POSSIBLY BE HELPFUL TO ME??

Linguistic analysis IS:

- Looking not just at *what* people say, but *how* they say it
- Paying attention to the "little words" (e.g., *you, might, not*)
- Noticing how we create politeness with language
- Thinking about the words that are *not* there

Linguistics: What Have You Done For Us Lately?

- Identifying different types of questions and responses in the operating room can help improve safety and communication (Bezemer et al., 2017)
- Providing concrete ways to talk about and teach patient communication (Maynard & Heritage, 2005)
- Analyzing high quality case presentations in order to help to teach it (Chan, 2015)
- Using certain pronouns can affect patient adherence (Falkenstein et al., 2016)
- Counting conversational turns between doctor and patient to predict patient satisfaction (Roter et al., 2008)
- Tracking clinician language to better understand how emotion and uncertainty may affect clinical reasoning (Konopasky et al., 2019)

Taking Turns in Conversation

HOW CAN WE ANALYZE TURN-TAKING?

1. Count **number** of turns
    - How do you define a turn?
2. Quantify **length** or **density** of turns
    - Length: number of words per turn
    - Density: number of statements or topics per turn
3. Characterize how people **take** and **yield** a turn
    - Taking: interruption, new topic, side comment, agreement
    - Yielding: questions, tag questions, trailing volume, pause, new topic

Taking Turns in Conversation

HOW CAN WE ANALYZE TURN-TAKING?

4. Characterize how people **support** turns
    - Back-channeling (“uh-huh,” “mm-hm,” “oh?,” “sure”)
    - Body language (nodding, smiling, open posture)
5. Note who **shifts the topic**
    - Who introduces a new topic?
    - How do they introduce the topic?
    - Who has the most knowledge about that newly introduced topic?

Taking Turns: Analysis

- Using the **turn-taking worksheet**, watch this next video and pick **one tool** (number of turns, length/density of turns, taking a turn, yielding a turn, supporting a turn, or shifting a topic) to use for analysis
- As you mark down the instances of your tool, think about how it might be used to support your program practically

Your Turn to Analyze Turn Taking

- Insert Video Clip Here

IMPRESSIONS/ THOUGHTS?

Saving “Face” and Hedging

WAYS PEOPLE MANAGE OPINIONS ABOUT OTHERS AND THEMSELVES

1. Kinda sorta maybe **hedging** your claims?
    - Qualifying degree (e.g., how much): “kind of,” “sort of,” “like,” “you know,” “a little,” etc.
2. Pointing out what is definitely **not** the case—**negation**
    - Comparing or alluding to what isn’t true: “You’re not being admitted to the hospital”
3. Noting what you **have to** do—**obligation**
    - Being compelled, more or less, to do something: “should,” “need to,” “got to,” “have to,” etc.

Saving “Face” and Hedging

WAYS PEOPLE MANAGE OPINIONS ABOUT OTHERS AND THEMSELVES

4. Shifting from “I” to “we” to people generally—**pronoun** use
    - First-person singular “I” shows individual involvement and agency
        - “I think you need to practice handoffs a bit more.”
    - First-person plural “we” shows joint action but can also sound condescending
        - “We need to get you some more practice with handoffs.”
    - Generic “you” or “they” indicates a general situation you have little control over
        - “You need to work on handoffs more in these kinds of cases.”
        - “They recommend practicing handoffs more in this situation.”

Saving Face & Hedging: Analysis

- Using the **saving face and hedging worksheet**, watch this video again and pick **one tool** (hedging, negation, words of obligation, pronouns) to use for analysis
- As you mark down the instances of your tool, think about how it might be used to support your program practically

Your Turn to Analyze Face Saving

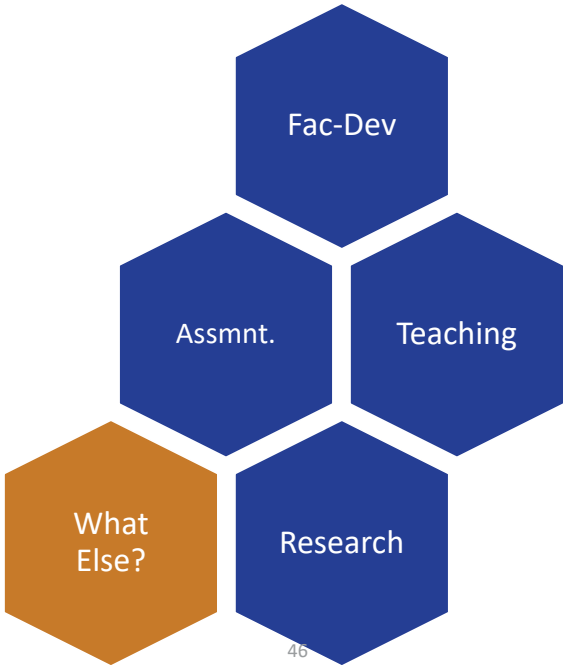
- Insert Video Clip Here

IMPRESSIONS/ THOUGHTS?

WRAPPING UP

Reflections and Looking Forward

How do you envision using at least one of these tool sets?



USU
Uniformed Services University

5 Tips

- Consent (consent handout)
- Managing your data (data management handout)
- Focusing your analysis
- Making time for planning, reading, analysis, etc. (readings handout)
- Setting reasonable expectations

FINAL QUESTIONS?

Thank You

- Alexis Battista, PhD, MBA
  Assistant Professor of Medicine
  Graduate Programs in Health Professions Education
  Uniformed Services University
  Email: alexis.battista.ctr@usuhs.edu

- Abigail Konopasky, PhD[2]
  Assistant Professor of Medicine
  Graduate Programs in Health Professions Education
  Uniformed Services University
  Email: Abigail.konopasky.ctr@usuhs.edu

USU
Uniformed Services University

References

Live Scenario - Gallstone Pancreatitis

Study Team Version

Leading Diagnosis:	Differential Diagnoses:	Contextual Factors:
Gallstone Pancreatitis	Top 2: Gallstone Pancreatitis Alcoholic Pancreatitis Others: Cholelithiasis Ascending Cholangitis Peptic Ulcer Disease Gallstones Alcoholic Hepatitis GERD	Diagnostic Suggestion

Setting: Emergency Department Consult

Scenario Length: 15 Minutes

Patient Name: Lt. Col. David (Dave) Turner

Age: 54

Date of Birth: November 13, 1964

Patient SP Characteristics Desired: No specific traits required.

Spouse: Anne Turner

Age: Early - Mid 50's, Attorney in private practice

Spouse Characteristics Desired: Similar in age but can be slightly older or younger.

ED Nurse or ED Physician: Study Team Member or Educator

SP Characteristics Desired: No specific traits required.

Case Stem Summary (see SP details and storyboard for in-depth information):

Lt. Col. (Air Force) David Turner is a 54 year old male [no restrictions on race/ethnicity] who presents to the emergency room with a 5-hour history of 8/10 mid-abdominal pain that radiates to his back and is associated with nausea and vomiting. The pain began while he was at a cookout and having drinks with his friends. Since then, he has been unable to eat or drink without exacerbating his symptoms.

He is accompanied by his wife, Anne Turner, who expresses concern [not nagging] about his recent drinking habits, which have become heavier over the past few months as his workload/caseload [he's an attorney] has increased. For example, she's noticed that along with his working more than usual, that he's also been going out for drinks at a bar with friends in the evenings. And, approximately two weeks before these symptoms began he was at a party with his Gulf War "military buddies" and he called her to pick him up as he was too drunk to drive home.

She is also concerned because alcoholism runs in his family and she thinks that alcohol is causing his current symptoms.

Standard Participant Details - Patient

Domain: Emergency Department

Leading Diagnosis: Gallstone Pancreatitis

Differential Diagnosis: Alcoholic pancreatitis, cholelithiasis, ascending cholangitis, peptic ulcer disease, gallstones, alcoholic hepatitis, GERD

Patient Name: Lt. Col. David Turner

Gender: Male

Age: Range 50-55 (Patient's age is 54)

Unique Findings/Characteristics: None

Setting: Emergency Department Consult

Timing: The scenario takes place at 2000 on a Saturday evening

General Appearance: Uncomfortable and nauseated but not writhing in pain.

Clothing/Gown: Already in a patient gown, in bed (ideally an ED type gurney)

Moulage: None

Chief Complaint: Abdominal Pain

Opening Statement by the Patient:

[David Turner]

" Like I told the other doctor [grimacing], I have a lot of pain in stomach here [gestures while circling an open hand over his upper abdomen] and it hasn't gone away like I'd hoped."

History of Present Illness: Mr. Turner presents to the emergency room on a Saturday evening with a 5-hour history of 8/10 mid-abdominal pain that radiates to his back and is associated with nausea and vomiting. The pain began while he was having some drinks and some food with his friends and since then has been unable to eat or drink without exacerbating his symptoms. His pain hasn't decreased and nothing seems to help, such as waiting for some time to pass, resting or finding a comfortable position. After trying these actions his wife called the nurse advice line (about 2 hours after symptom onset) and they recommended he come to the ED.

After arriving in the ED, Mr. Turner was seen in triage but the ED is busy and it took about an hour for him to be taken back. When we was taken back, he was seen by the ED physician who

then ordered blood drawn for labs, an EKG [showing Sinus Tachycardia] and he's supposed to go over an have an ultrasound soon [Radiology is also backed up due to the business of the ED]. He has an IV started [1L NS] for fluid resuscitation and was given Ketorolac [Toradol] 30 MG IV for pain. Since starting the IV fluids and receiving the Ketorolac his pain has improved to 6/10.

How long has *this* episode lasted: The current episode has lasted for about 5 hours.

Affect and behavioral expectations: He is uncomfortable and feels awful.

Onset of this complaint: This episode came on when he have a few beers and grilling hamburgers with a couple of friends. At this cookout, he had a hamburger, some potato chips, and two beers.

Location of Pain: Mid-upper abdomen - and radiating into his mid-back area

Character/Quality of this Episode of Pain: The pain is "deep and squeezing" in nature [gestures over the upper-abdomen by squeezing his hand shut].

Severity/Pain Scale: At onset the pain was a 6/10 and over the next hour it increased to an 8/10 [leading him to seek care in the ED]. The pain was an 8/10 when he arrived at the ED but has declined to a 6/10 after receiving some IV fluids and pain medication.

Frequency of Pain: This pain hasn't gone away since it started 5 hours ago.

Aggravating Factors: Any eating or drinking or lots of movement.

Alleviating Factors: Avoiding food and liquids and is reluctant to move around a lot.

Similar Episodes: He had a single similar episode of pain - with similar type and location of pain [mid-abdomen - but not radiating to the back] that he recalls being about a 3-4/10 [pain] about 4-5 weeks ago. He was out having a few drinks with his friends after work. He had some nausea at the time and figured he just needed to hold off on anymore drinks that evening [he had already had 3 beers and a club sandwich]. The episode that occurred 4-5 weeks ago resolved on its own and has not returned so he hasn't sought care for it because it resolved.

Last Meal: The pain came on in the late afternoon while he was hanging out with a few friends grilling some hamburgers. He also had two beers before the pain onset.

Other Symptoms Associated with Current Chief Complaint: Nauseated, Vomited 3 times. No blood in vomit. No diarrhea or constipation.

Past Medical History: Diabetes Mellitus and hyperlipidemia - both diagnosed about a year and a half ago.

Current Medications:

- Atorvastatin (Lipitor) 40 mg every day at bedtime for hyperlipidemia, last dose was last night
- Metformin 500 mg twice daily with breakfast and dinner (with a meal), last dose was this morning with breakfast.

Surgical History: No prior surgeries.

Social History:

Occupation: Active Duty Attorney [JAG], works at the Pentagon

Religion: Does not regularly attend church and doesn't identify strongly with any one particular religion.

Family: Married 27 years, 2 children (son and daughter) - both in their early 20s. Daughter is thinking about law school. Son is considering applying to medical school (conveys the family is very educated and house is likely full of stress.

Lifestyle: Over the past 3-4 months, Lt. Col. Turner's workload and caseload has increased significantly. This increased workload has led him to work longer hours and he's had little time to engage in the things he typically enjoys, such as playing basketball, and refurbishing muscle cars (the current vehicle he is refurbishing is a 1968 Dodge Charger that he takes to local "car cruise ins" [low - key informal gatherings of individuals who refurbish muscle cars where they show off their cars and their work]. Much of his work is classified and he's not able to discuss his cases or the matter of law that is involved. Since this increased workload he's also engaged in more regular bouts of social drinking [3-4 times per week] where he goes out with a couple of work colleagues after work to have a few drinks together.

Habits:

Alcohol: Typically [when not under great stress] he has either a beer or a glass of scotch daily after work. However, recently [in the last 3-4 months] his intake has increased to 2 - 3 drinks [beer or scotch] a day. He goes out with colleagues/buddies about 4 days a week and may occasionally go out solo. Approximately 4-5 weeks ago he attended a party with his Gulf War "military buddies" and had to call his wife for a ride home because he had had too much to drink. His wife has become increasingly concerned about this change - and his decline in his usual hobbies - especially after his drinking so much he needed a ride home. Her increasing concern has recently become annoying for him. CAGE =1/4.

Tobacco: Denies

Diet: He's trying to watch his diet and eat lower fat foods since being diagnosed with hyperlipidemia, but he still eats on the run a fair amount (often sub shops, the occasional hamburger), he typically has 2-3 cups of coffee in the mornings.

Drug Use: Denies

Exercise: Doesn't go to a gym all that often, relies primarily on walks with his wife and basketball with his friends. He's always been able to pass his PT test without a lot of training.

Sleep habits: He hasn't been sleeping well due to recent workload/stress, typically getting 5-6 hours of sleep a night. This isn't too far from off from his usual amount of sleep, however.

Family History:

- Father – Living. All he really knows about his dad's history is his alcoholism and that he's now sober. Also had an episode of peptic ulcer disease in his mid-50s.
- Mother - Living; medical history includes depression (well controlled by taking a low dose antidepressant [unsure of exactly which antidepressant] and intermittent counseling) and hyperlipidemia.
- Siblings - Younger brother [49] who is an auto mechanic who also served in Gulf War. He has hypertension controlled on medications.

Physical Exam Results:

Extremities:

- CNS: Alert, oriented
- HEENT: Dry mucous membranes
- Cardiovascular: Tachycardic, regular rhythm. No murmurs, rubs, or gallops.
- Pulmonary: Clear to auscultation bilaterally
- Abdomen:
    - Decreased bowel sounds.
    - Tenderness to palpation in epigastrum without guarding or rebound.
    - No hepatosplenomegaly
    - Murphy's sign: negative
    - Any swelling in the abdomen: No swelling
- Extremities: No cyanosis, clubbing, edema
- Musculoskeletal: Normal
- Integumentary: Normal
- Psychological: Normal

Vitals:

- Pulse: 106
- BP: 150/82 [If asked the SP can share that normally his blood pressure is around 130/80]
- Respirations: 16
- SP02: 98% RA
- Temp: 99.6

12 Lead: Sinus Tachycardia

Labs:

ED Lab Results:

AST	94
ALT	89
Bilirubin	1.2
Amylase	490
Lipase	4700

CBC	
WBC	9.8 (68% neutrophils, 24% lymphocytes, 5% monocytes, 2% eosinophils, 1% basophils)
Hemoglobin	12.5/38 crit
Platelets	415,000

BMP	
Sodium	140
Potassium	3.8
Chloride	104
Bicarb	26
Glucose	98
Creatinine	1.1
BUN	18

Storyboard

0-2 min **Receive ED Consult Request** 	1. After the participant receives their instructions and indicates they have no further questions they will receive a call from the ED physician [performed by an educator or study team member] requesting a consult. This is the start of the live scenario. [ED Physician] Hey, we've got a 52 year old man with a history of pretty significant alcohol use here with abdominal pain...his exam is pretty benign other than moderate epigastric tenderness without peritoneal signs. I've sent some labs but they haven't resulted yet and I've ordered an abdominal ultrasound but radiology is pretty backed up. His EKG shows sinus tach but is otherwise unremarkable. He's not tolerating anything by mouth and needs better pain control and fluid resuscitation, so he needs to come in. We're really busy and would appreciate if you could come see him now. 2. The scenario will run in real time rather than speeding time up. 3. Participants will only have access to those diagnostic findings that they seek and perform. Diagnostic findings will not be volunteered without prompting because this will allow us to assess their reasoning processes, including, what they prioritize early in their assessment, what diagnostic evidence they use to make choices, and the order in which they ask for and use it.

Time	**Settings and Changes**	**Patient Responses/Cues**	**Potential Participant Actions**

2-15 min **Initial Interview** 	**Vitals:** Pulse: 106 BP: 130/82 Respirations: 16 SP02: 98% RA Temp: 99.6 Cardiac rhythm – Sinus tachycardia Lungs – clear bilateral	[Pending]	1. Interview & Assess the patient. 2. Review existing ED record.
@ minute 12 **Lab Results Arrive**	Unchanged	SP nurse hands the participant the labs that have just come back.	1. Review Lab Findings
12 - 15 min **Scenario End -** 	Unchanged	Patient and wife ask the participant what they think is going on? After or while the physician participant shares their summary of what they think is going Mrs. Turner raises her concern about alcohol involvement.	1. Discuss next steps with the patient and his wife. 2. Seek a consultation with another specialist. 3. Anything else?

Scenario Context Mapping Questionnaire

Question	Free Text Response
What are the typical physical tools (e.g., stethoscope, ultrasound device, etc.) that are needed or found in this type of scenario?	1. Exam Room as if in the emergency department. 2. Patient wearing a gown 3. Simulated Stethoscope - set to decreased bowel sounds. 4. Sp02 Probe on continuous 5. BP Cuff or automated device on and set to assess pulse, 02 sats, blood pressure q15. 6. Paper medical record to share with the participant in advance of their entering the patient room.
What props would further support the clinical situation (e.g., standard patient with moulage on the [location], human patient simulator, etc.)?	1. Standardized patient to portray the patient. 2. Standardized participant to portray the patient's spouse. 3. Standardized participant, educator or study team member to portray the ED physician to support the hand-off at the start of the simulation. 4. Educator or study team member to portray an ED nurse [will help keep the flow of information going and address any issues that arise that are unexpected]? 5. Moulage - none
What personal patient safety equipment should be available for the scenario?	1. Gloves 2. Alcohol gel
What are the diagnostic findings that would be needed to support participants as they make or confirm a diagnosis in this scenario?	Findings available prior to physician entering the room: ● Today's vital signs with trends over past hour and triage vitals on ED record. ● Labs this visit - arrives after study participant interviews the patient. ● Ultrasound ordered but awaiting to go over to radiology. ● EKG - Showing sinus tachycardia at 108

	● What else would the ED nurse and physician done so far? How would this be recorded? The triage note, EKG, and any other paper records would be in a binder in the workroom.
What diagnostic activities (e.g. auscultation, palpation, etc.) would normally be used in this type of scenario? By who?	What would the ED nurse have done? 1. Started Fluid Resuscitation - 2L NS hanging at 200cc/hr. 2. Given pain management [Ketorolac 30 mg IVP] What would the ED physician done so far? 1. Physical exam (focusing on abdomen, pulmonary and cardiac). 2. EKG 3. Ordered Ultrasound 4. Ordered Fluid Resuscitation - 2L NS hanging at 200cc/hr. 5. Ordered pain management [Ketorolac 30 mg IVP] What do we think the participant will do? 1. Visualization (Abdomen) 2. Auscultation (Cardiac, Pulmonary, Abdomen) 3. Palpation (Abdomen, Percussion) 4. Diagnostic questioning
What types of therapies (fluid challenge, medications, etc.) would typically be offered in this type of scenario?	1. Fluid Resuscitation - 2L NS hanging at 200cc/hr. 2. IV Ketorolac 30 mg IVP

Question	Free Text Response
What rules would normally guide or govern care or behavior in this scenario? Rules –are conventions or guidelines that regulate activities. For example, is there a standard that governs how all patients who present with trauma are assessed or treated? Are there standard order sets that govern care?	ED Assessment Guidelines: Admission Guidelines:
Who is typically present during a scenario such as this, and what role do they play during the event? Please give a brief description of what the role entails. Please include all roles typically present. Roles – are the division of labor. For example, so you may have a physician stationed at the head of the bed whose role is airway. The person present is the physician and their role is airway management.	1. 1 Physician (Study Participant) 2. 1 standard participant as the patient. 3. 1 standard participant as the spouse. 4. 1 standard participant as the ED nurse who can also make the call to the participant.

Author(s):
Steven Durning, MD, PhD, FACP
Robert Trowbridge, MD
Alexis Battista, PhD

SME Consults/Authors
Abigail Konopasky, PhD
Megan Ohmer, BS
Sarah Krajnik, RN, BSN
Anna Howle, MFA
Jeff Mikita, MD
Richard Wu, MD

Live Scenario - Stable Angina
Simulated or Clinic Setting

Study Team Version

Primary Diagnosis:	Alternative Diagnosis Options:	Contextual Factors:
Stable Angina	Unstable Angina Acute Myocardial Infarction Pulmonary Embolism Aortic Stenosis GERD Peptic Ulcer Disease Costochondritis	-EHR unavailable -Patient suggests alternative diagnosis (at least 2)

Setting: Outpatient Clinic

Name: COL Paul Stone (JAG Corps)

Age: late 40's or early 50's

Date of Birth: April 2, 1965

SP Characteristics Desired:

Male, normal to slightly overweight, no preference on ethnic background but we remain consistent once we've cast this role.

Summary of Case Stem (see SP details and storyboard for in-depth information):

COL Stone is a late 40's or early 50's male who has been experiencing intermittent episodes of burning and pressure in his chest. He also feels winded when this happens (though he doesn't make this connection). He insists that it's probably just bad heartburn, like his mom has, because he's worried that this could be more serious condition. He's been reading about possible causes of chest pressure and burning. The scenario begins when the physician enters the outpatient clinic room to see the patient. The patient is not having any pain at this time.

Standard Participant Details

Domain: Outpatient Clinic

Diagnosis: Stable Angina

Alternative Diagnosis: Unstable Angina, Acute Myocardial Infarction, GERD, Peptic Ulcer Disease, Pulmonary Embolism, Aortic Stenosis, Costochondritis

Contextual Factors:

- No EHR - the system is down at the time of the visit.
- The patient insists on at least two alternative diagnoses (e.g., **heartburn (GERD) "like his mom"**, pulmonary embolism because of recent travel to San Antonio, Anxiety because he's been under a lot of stress and read about anxiety as a possible cause)

Gender: Male

Age: Range 48-55

Unique Findings/Characteristics: None

Setting: Outpatient Clinic Exam Room

General Appearance: Normal, well kept, without distress

Clothing/Gown: Street clothes.

Moulage: None

Chief Complaint:
Intermittent tightening and burning sensation in the chest for past 3 weeks that won't go away.

Opening Statement by the Patient:
SP: "I've been getting a burning pain right here doc..." [gestures with open hand over center of chest using a circling motion to indicate that the pain is generally felt to the center of the chest]

SP: I...I don't know what it is but it seems like I get winded more easily, too. I figured it would go away on its own, but it hasn't and my family keeps insisting that I come and see you.

History of Present Illness:
Presents with complaint of intermittent burning and pressure in the chest for past few weeks, the sensations occurs with exertion (e.g., running, strenuous exercise). He also experiences shortness of breath or a sensation of feeling "winded" at the same time. Walking and short flights of stairs are fine.

How long do the episodes last:
The pain/burning episodes can last several minutes and get worse if he continues strenuous physical activity. The sensations improve within a few minutes of stopping/resting.

Affect and Behavioral Expectations:
Cooperative, good historian, concerned. When his symptoms started he started taking Tums and Prilosec because he thought it might be GERD - something his mom has had. However, recently, he is aware that this could be something serious because he's been reading about potential causes of the pain, such as heart attack. He tries to deny this by suggesting at least 2 alternative diagnosis, such as a pulmonary embolism or deep vein thrombosis (he's recently flown to San Antonio).

Onset of complaint: Approximately 3 weeks ago

Location of Pain:
Tightening or burning sensation in his chest – uses full hand to gesture to center area of chest consistent with angina.

Character/Quality of Pain/Complaint: Feels like a burning or tightening sensation.

Severity/Pain Scale: 0 (currently); 4-5 (with symptoms)

Frequency of Pain: Anytime he exerts himself - at least a few times a week.

Aggravating Factors: Exercise or more activity than normal walking. He's been more active because his PT test is coming up and his workload has been heavier than usual. Because of his heavier workload he'd not been able to keep up with his regular PT.

Alleviating Factors: Rest. Subsides within a few minute of resting. The pain is not relieved with Tums or Prilosec.

Similar Episodes: None

Current Medications:

- HCTZ 25 MG once daily,
- Metformin 500 mg PO, Daily,
- Tums - started taking when he first had the symptoms assuming it was heartburn.
- Prilosec - started taking when he first had the symptoms assuming it was heartburn.

Past Medical History: High Blood Pressure & Type 2 Diabetes - onset x 4 years ago

Surgical History: Appendectomy at age 25

Social History:

Occupation: Active Duty Lawyer (JAG)
Religion: Catholic
Family: Married 25 years, 2 children - both high school aged.
Lifestyle: Busy workdays - case-load has been pretty heavy lately and he'd gotten off track from his usual PT schedule. He's been trying to get back on schedule recently until the chest pressure/burning and shortness of breath started.

Additional History:

- Recently flew to San Antonio to attend a funeral for a former colleague.
    - The goal of inserting this history is to give the participants added concerns about a Pulmonary Embolism or a Deep Vein Thrombosis that can occur with long sitting periods, such as flying. This isn't something to volunteer - but if the participant asks about it - just say it very frankly or matter of fact.

Habits:

Alcohol: occasional drinker – 4-5 drinks per week. CAGE =0/4
Tobacco: History of smoking, 1 ppd, 20 years, quit prior to surgery
Diet: High fat (eating out/fast food) and 4-5 cups of coffee (higher caffeine)
Drugs: Denies
Exercise: - Gotten off track from his usual PT schedule due to heavy case-load. Had been trying to get back on track because his PT test was coming up but started experiencing chest pressure/burning and shortness of breath.

Family History:

- Father – alive, unspecified heart problems - he isn't very forthcoming when it comes to discussing medical issues.
- Mother – alive, frequent heartburn/GERD, diabetes
- Siblings - no siblings.

Physical Exam Results:

- CNS: Alert, oriented
- HEENT: Normal
- Cardiovascular: regular rhythm, no murmurs, rubs, gallops
- Pulmonary: Clear to auscultation
- Gastrointestinal: normal
- Musculoskeletal: Normal
- Integumentary: Normal
- Psychological: Normal

Vitals:

- Pulse: 78
- BP: 156/87
- Respirations: 18
- SP02: 98% RA
- Temp: 98.5

12 Lead:

- Current 12 Lead - Normal

Labs:

- Preoperative Labs within 6 months:
    - Hemoglobin = 10 gm/dL
    - Basic Metabolic Panel = normal
    - A1C=6.5
    - Liver Panel
    - CBC
    - HIV - Negative

SP Statements (do not have to adhere verbatim - but suggested)

Opening statements:

Col. Stone: "I've been getting a burning pain right here..." [gestures with open hand over center of chest using a circling motion to indicate that the pain is generally felt to the center of the chest]

Col. Stone: "I...I don't know what it is but it seems like I get winded more easily, too. I figured it would go away on its own, but it hasn't and my family keeps insisting that I come and see you."

Contextual Factors Statements:

"I came in my family keeps insisting on it--it's probably nothing--maybe acid reflux or something like that like my mom."

"When I exercise, I feel this tightening--I'm thinking maybe I have an anxiety disorder or something like that because I'm swamped at work."

"I feel like I'm a certified doctor now--I browse WebdoctorMD and I feel like I can diagnose myself--I'm thinking maybe this could be some kind of pulmonary em-embolism or something like that. Maybe deep vein thrombosis. I'm just saying--I don't know. But I-I think it's just GERD."

Asked about the Prilosec and Tums: "It kind of helps I guess, a little bit. In my mind's eye, I guess." [but if asked about whether it helps when he exercises: "No, no, I don't think so."]

Additional statements that may be used regarding symptoms - if asked:

Asked when he noticed it started: "I've got my PT test coming up and I've been increasing my activity and that's when I started noticing it."

Asked about activity: "I do fast walking, go to the gym, the elliptical--I mean I'd like to do a little running, but I haven't gotten into that yet."

Asked to compare this to past pain: "In the past I've gotten a little winded, sweated a little bit, but never this burning--I mean I know I'm not 25 anymore, but I've never experienced that before."

Asked about exercise routines: "I'm a JAG member, so my caseload is heavy and, you know, long hours, you know...it isn't conducive to a healthy lifestyle. So I'm out of shape, I've put a few pounds on."

Asked about diet: "I have been popping the Tums. You know, I have long hours, and I don't know if you've seen recently at the Pentagon--they have a lot of foods there, not many of them healthy that much."

Asked about duration of pain: "It lasts for, like, several minutes and then, like, if I stop, in a few minutes it goes away and I figure like if I stop, stop for today, and then come back tomorrow I'll be able to go a little bit longer--you know, progressive training and stuff."

Asked about BP: "I go in every 3 months and they monitor it there--they say it seems to be pretty well controlled."

Asked about family medical history: "My father has this--I guess you'd consider it some kind of unspecified heart problem situation--because he's of that generation where you don't complain. I just know there is a heart issue of some kind. I'm sure my mother knows. She probably wouldn't tell me either, but she's a bit more forthcoming. She suffers from acid reflux. And she has Type II diabetes."

Asked about travel: "A colleague of mine and I had to fly out to San Antonio a few weeks ago."

Proposed Storyboard	
0-3 min **Hand -Off and Scenario Start** 	1. After the participant receives their instructions and indicates they have no further questions (see participant hand off script below) they will receive details about their patient in the form of doorway information with patient details, i.e., reason for today's visit, vitals, past records (6 mos). ***This will be given to them by a study coordinator using a standardized document.** 2. The participant will have time to review this document and then will determine when they are ready to enter the clinic exam room where the SP will be waiting (in street clothes). This is the start of the live scenario. 3. Note - the scenario will run in real time rather than speeding the time lapses up. We reason that by running the scenario in actual time will more accurately allow us to assess the participant's clinical reasoning processes. Speeding time up sets up a situation in which we could skew participant's reasoning by making time move more quickly than typically experienced in the actual clinical setting. 4. Note - participants will only have access to those diagnostic findings that they seek and perform. Diagnostic findings will not be volunteered without prompting because this will allow us to assess their reasoning processes, including, what they prioritize early in their assessment, what diagnostic evidence they use to make choices, and the order in which they ask for and use it.

Time	Manikin Settings and Changes	Patient Responses/Cues	Potential Participant Actions
1-10 min **Initial Interview** 	**Vitals:** Pulse: 76 BP: 156/87 Respirations: 18 SP02: 98% RA Temp: 98.5 Cardiac rhythm – Sinus Lungs – clear bilateral	**Affect:** Appears concerned and somewhat anxious. **SP Opening Statement:** Follows after the SP and physician exchange greetings and the physician asks how he is doing or what brought him to the clinic today. SP: "I've been getting some burning pain right here doc..." [gestures with open hand over center of chest using a circling motion to indicate that the pain is generally felt to the center of the chest] SP: I...I don't know what it is but it seems like I get winded more easily. I had hoped it would go away on its own, but it hasn't and it's got me worried that maybe this is a problem from my knee surgery or maybe one of those medications." [expresses with moderate level of concern]	1. Participants should conduct an initial interview. 2. The interview should give way to a physical exam.
5-15 min **Physical Exam** 	**Vitals:** Pulse: 74 BP: 156/87 Respirations: 18 SP02: 98% RA Temp: 98.5 Cardiac rhythm – Sinus	**Affect:** Appears concerned and somewhat anxious. **SP Statements:**	1. Reviews/analyzes vital signs 2. Continues communication with the patient (diagnostic questions,

	Lungs – clear bilateral	As the physician conducts the physical exam interrupt by suggesting an alternative diagnosis such as, “now that you mentioned my mom, maybe this is just some really bad heartburn”	education and counseling) 3. May review findings as they become available (e.g., 12 Lead)
15 min **Scenario End -** 	Scenario End	Scenario End	1. Facilitator should end the scenario if the participant hasn’t done so yet by entering the room and announcing that we will stop here.

Scenario Context Mapping Questionnaire

Question	Free Text Response
What are the typical physical tools (e.g., stethoscope, ultrasound device, etc.) that are needed or found in this type of scenario?	1. Physical Exam Room as if in a clinic setting. 2. Patient Gown (on standby) 3. Stethoscope 4. Sp02 Probe 5. BP Cuff or automated device to assess pulse, 02 sats, blood pressure. 6. 12 lead machine or replica available but not in the room. 7. Clock with second hand on wall 8. Paper medical record to share with the participant in advance of their entering the patient room. 9. Diagnostic data from other prior medical exams that are age appropriate - available for the participant to view if asked for.
What props would further support the clinical situation (e.g., standard patient with moulage on the [location], human patient simulator, etc.)?	1. Standard patient to portray the patient. 2. Moulage an older scar commonly associated with total knee (long incision over the knee) - well healed.
What personal patient safety equipment should be available for the scenario?	1. Gloves 2. Alcohol gel
What are the diagnostic findings that would be needed to support participants as they make or confirm a diagnosis in this scenario?	Findings available prior to physician entering the room: ● Today's vital signs ● Labs from most previous visit ● Summary indicating current on vaccines, most recent HIV test negative, most recent well women visit normal approx. 6 months ago. Findings available during simulation ● 12 Lead EKG - normal Findings not available even if requested*: ● CXR

	• Labs • Referral for holter monitor • Referral for stress test
What diagnostic activities (e.g. auscultation, palpation, etc.) would normally be used in this type of scenario?	1. Visualization 2. Auscultation 3. Palpation 4. Obtain blood pressure 5. Diagnostic questioning
What types of therapies (fluid challenge, medications, etc.) would typically be offered in this type of scenario?	1. Physical exam 2. Social and emotional support 3. Education and counseling 4. Zantac 5. Aspirin 6. NTG 7. Oxygen

*These findings aren't available because we are running the scenario in live time. Although the participant may request additional tests, they wouldn't normally become immediately available.

Question	Free Text Response
What rules would normally guide or govern care or behavior in this scenario? <u>Rules</u> –are conventions or guidelines that regulate activities. For example, is there a standard that governs how all patients who present with trauma are assessed or treated? Are there standard order sets that govern care?	Assessment Guidelines: 1. AHA/ACC Stable Angina Guidelines. 2. SOP for WRNMMC Clinical Operations Guidelines 3. Stable angina guidelines from other participating institutions if we collect data off site.
Who is typically present during a scenario such as this, and what role do they play during the event? Please give a brief description of what the role entails. Please include all roles typically present. <u>Roles</u> – are the division of labor. For example, so you may have a physician stationed at the head of the bed whose role is airway. The person present is the physician and their role is airway management.	1. 1 Physician (Participant) 2. 1 Standard patient (Patient)

Author(s):
Steven Durning, MD, PhD, FACP
[Add Others]

Adaptation Authors:
Alexis Battista, PhD
Steven Durning, MD, PhD, FACP
Sarah Krajnik, RN, BSN
Anna Howle, MFA
Jeff Mikita, MD
Jeff La-Rochelle, MD
Paul Hemmer, MD, MPH

Live Scenario – Inpatient Trauma[1]

Summary:
This scenario is set in the inpatient setting where Mr. Carl Fisher, a 50-year-old male, was admitted the previous evening (around 0500) with a diagnosis of cellulitis. Today (time of the study) Mr. Fisher has tripped and fallen on the sink (right chest) in his bathroom which causes a rib fracture and subsequent pneumothorax. The fall was not witnessed, but he told the nurse that he got tangled up in the wires and feet of his IV pole and fell. After his fall, the nurse responded, helped him settle back to bed, asked the technician to check his vitals and called the physician to come check him out. At the beginning Mr. Fischer isn't in significant distress; however, as the scenario progresses the pneumothorax will progress to a tension pneumothorax that requires treatment with needle decompression or placement of a chest tube. Study participants are also expected to eventually call for a rapid response.

Leading Diagnosis	Differential Diagnoses	Contextual Factors
❖ Pneumothorax → Tension Pneumothorax	❖ Rib Fracture ❖ Hemothorax ❖ Pulmonary Contusion ❖ Pulmonary Embolism ❖ Syncope related to cardiac arrhythmia (e.g., a fib, Vtach) ❖ Hypovolemia (e.g., dehydration, sepsis) ❖ Anaphylaxis ❖ Stroke/CVA ❖ Pneumothorax ❖ Cardiac Contusion ❖ Splenic or liver laceration/hematoma	❖ *Limited knowledge of the patient* ❖ *Emotional volatility due to increasing hypoxia* ❖ *Increasing acuity of presentation* ❖ *Team-based clinical reasoning*

[1] Author(s):
Alexis Battista, PhD
Steven Durning, MD, PhD
Anna Howle, MAC
Sarah Krajnik, RN, BSN
Jeff La-Rochelle, MD, MPH
Paul Hemmer, MD, MPH
Jeff Mikita, MD
Walter Kucera, MD
Matthew Nealeigh, DO
Sarah Ordway, MD
Thomas Mellor, MD
Megan Ohmer, BS

Last Revised 08 May, 2018

Scenario Storyboard			
Scenario Start  **0-3 min**	1. After the participant receives their instructions and expectations for participants and indicates they have no further questions (see participant hand off script below) they will receive a phone call (using a study team designated phone) from the participant portraying the patient's primary nurse. 2. **Opening Statement** [Ideally read verbatim for continuity by SP Primary Nurse]: *"Hi, this is [name of nurse] up on 4 Center. I'm calling about Mr. Fisher – he was just admitted from the ED about 30 minutes ago with cellulitis. We got his antibiotics hung and he got up to go to the bathroom and fell. He says he got tangled up in his IV pole. He's alert and oriented, denies any loss of consciousness and says he didn't hit his head...but he's got some pain on his chest where he hit the sink."* His vital signs are: *Pulse: 98* *BP: 130/80* *Respirations: 18* *SP02: 98% RA* *Pain: 7/10 at the injury site* 3. The study participant can ask further questions of the primary nurse, such as how is he doing now? What are his vital signs? Can you get a 12 lead so I can see it when I get there? If the participant asks for a 12 Lead let the participant know you will work on getting one. See details for Primary Nurse for additional responses. 4. Following the call - the study coordinator will direct the participant to the patient's inpatient room.[2]		
Time	**Goals and monitor settings**	**Patient Responses/Cues**	**Potential Participant Actions**

Commented [1]: +anthony.artino@usuhs.edu Tony, This is the scenario storyboard for the new inpatient trauma scenario for you to review so you can get a better understanding of what participants will be expected to do as this scenario unfolds. I will share the specific questions that we've designed so far under separate cover. I also sent a link to Tim.
Assigned to Anthony Artino

Commented [2]: +tc454@rci.rutgers.edu Tim, This is the scenario storyboard for the new inpatient trauma scenario for you to review so you can get a better understanding of what participants will be expected to do as this scenario unfolds. I will share the specific questions that we've designed so far under separate cover.
Assigned to tc454

Commented [3]: This could be a good opportunity to study their forethought processes and should be recorded and re-played as a part of the think aloud.

[2] The scenario will run in real time rather than speeding the time lapses up. We reason that by running the scenario in actual time will more accurately allow us to assess the participant's clinical reasoning processes. Speeding time up sets up a situation in which we could skew participant's reasoning by making time move more quickly than typically experienced in the actual clinical setting. Participants will only have access to those diagnostic findings that they seek and perform. Diagnostic findings will not be volunteered without prompting because this will allow us to assess their reasoning processes, including, what they prioritize early in their assessment, what diagnostic evidence they use to make choices, and the order in which they ask for and use it.

Physician arrival at bedside.  **3-8 min**	Goal: During this segment the patient will compensate to give the participant an opportunity to assess the patient and consider potential differential diagnoses. Vital Signs: *Pulse: 106* *BP: 124/78* *Respirations: 20* *SP02: 97% RA* **For the primary nurse SP – for vital signs checks in this stage of the scenario please take the cuff and SPO2 probe off the patient.	1. Patient is in the patient bed with staff having already obtained a set of vital signs at the participants' arrival. 2. In pain, cooperative, appears uncomfortable – braces *(holds area with arm or hand type gesture*) his injured side and tries to minimize movement (*stiff - guarded movement*), respiratory effort is mildly elevated and not deep because it's painful to take in a full breath. - Pain without movement is achy and 4/10 -- but "*fears*" the sharper pain that comes with movement. - Pain worsens - sharp - with movement or deep breath (8/10). - Increased pain and tenderness to the right lateral chest with palpation (*SP will guard if palpation attempted*) (8/10).	1. Communication with the patient (introduction, diagnostic questions). 2. Conduct an initial physical exam (focused). 3. Reviews/analyzes initial set of vital signs. 4. May request repeat vital signs. 5. Request and review patient admission documents, diagnoses, prior dx, and medications. 6. May request supplemental Oxygen, 7. May request pain management (e.g., Tylenol, Tylenol with a narcotic PO or medication by IV). 8. May request an ultrasound machine.
Time	**Goals and monitor settings**	**SP Responses/Cues**	**Potential Participant Actions**
Continued assessment,	Goal:	1. Pain may be decreased (5 or 6/10) if analgesia	1. Reviews/analyzes vital signs

Commented [4]: This segment will allow us to examine how they gather information from the patient (verbally and gesturally), the primary nurse and the medical record-- and the PEF/TA will help contextualize their reasoning.

Commented [5]: This scenario will also allow us to study how their differential diagnoses evolve from call to initial assessment and again during increased acuity.

patient deteriorates and rising acuity  **8 - 18 min**	During this time frame the patient will start to deteriorate which will further introduce increased acuity and increase patient anxiety. Vitals (trending down to the following): Pulse: 130 BP: 86/60 Respirations: 26 SP02: 91% with supplemental Oxygen. *Decreased breath sounds on the right (injury site). * *For the primary nurse SP – for vital signs checks in this stage of the scenario please leave the cuff and SPO2 probe on the patient.	previously given (e.g., Morphine). 2. Difficulty breathing continues to worsen even with supplemental oxygen. 3. Anxiety continues to the point where the patient occasionally is frustrated with the medical providers, including the study participant. 4. If the participant doesn't recognize the changes in patient affect and vital signs, the SP can elevate their level of frustration with the study participant. 5. Near the end of this stage the patient will start to become somnolent. 6. At the point of somnolence and if no RRT team called yet the SP nurse will cue the participant to call an RRT.	2. Continues communication with the patient. 3. May request supplemental Oxygen (e.g., nasal cannula, ox mask, simple mask, NRB). 4. May request subsequent vital signs checks (at this stage of the scenario the primary nurse will leave the BP cuff on) 5. May request moving crash cart into the room along with continuous 3 - lead ECG monitoring with a monitor 6. May request a stat chest x ray 7. May request an ultrasound machine. 8. May request additional vascular access be started 9. May choose to call rapid response or senior resident. 10. May decide to treat the tension pneumothorax and perform a need decompression or a chest tube.
Time	**Goals and monitor settings**	**Patient Responses/Cues**	**Potential Participant Actions**
Option 1: Participant chooses to support &	Goal: During this time frame the first of the rapid response teams will arrive	1. In the event the participant does not know how or feel comfortable performing a needle	1. The participant will give handoff report to the arriving RRT team member.

Commented [8]: This would be a good opportunity to insert a series of microanalysis questions that the primary nurse or the ICU nurse asks, such as "what are you thinking is going on?"

await RRT Arrival  **XX - 20 min**	and seek an initial report on the patient's status. The goal will be to allow this reporting to be completed because it represents an opportunity to learn about their understanding of the situation. After report the study team will stop the scenario. Vitals (trending depends on actions): Pulse: 136 BP: 84/60 Respirations: 26 SP02: 90% with supplemental Oxygen.	decompression or placement of a chest tube (even after cueing) the patient's condition will continue to deteriorate. 2. The RRT team will focus on getting report.	
Time	**Goals and monitor settings**	**Patient Responses/Cues**	**Potential Participant Actions**
Option 2 & 3: Needle decompress/Chest tube & await RRT Arrival to Handoff to	Goal: During this time frame if the participant chooses to treat the tension pneumothorax the	1. If they request a chest tube tray or a needle for decompression the primary nurse in the scenario will respond:	1. The participant needle decompresses or places a chest tube. 2. The participant may request a RRT – the primary nurse will cue this if not called and

Commented [6]: This will be the equivalent to an unprompted think aloud when they give hand off – similar to their talking with the patient about possible diagnosis in the other scenarios; however, here they will be talking with a fellow healthcare professional so that might be different.

Commented [7]: It might also give us a better idea of cognitive load because, in speaking with another health care professional, you'd expect the participant to use advanced "semantic qualifiers" and if s/he doesn't, that could indicate higher load.

End the Scenario  **XX – 20ish min**	patient will begin to improve. Shortly thereafter, the first of the rapid response team members will arrive and seek an initial report. Following report, the study team will end the scenario. Vitals (trending upwards): *Pulse: 120* *BP: 94/70* *Respirations: 22* *SP02: 94% with supplemental Oxygen.*	*"I'm going to call an RRT"* (if they don't call one) *"Let me get the crash cart"* (if they call for a chest tube/needle) 2. If the participant decompresses or places a chest tube the patient will experience pain during the procedure but will experience relief after placement.	request to do a decompression/chest tube. 3. May request regular vital signs assessments/updates. 4. The participant will give handoff report to the arriving RRT team member to allow for an opportunity to learn about the participant's understanding of the situation.

Simulated Activity System Questionnaire

Tools and Structured Interventions

Question	Free Text Response
What are the typical physical tools (e.g., stethoscope, ultrasound device, etc.) that are needed or found in this type of scenario?	For the patient room: 1. Inpatient med/surg room (med/surg ward) 2. Inpatient bed (med/surg ward) 3. IV pole x 1 (the one he tripped on) 4. Bedside table 5. Headwall with air/oxygen ports -simulated 6. Clock 7. Fresh linens and 2 gowns for SP 8. BP Cuff/Sp02 probe with automated device that can assess pulse, pulse ox, blood pressure with pre-printed & laminated vital signs ready/stored for easy access. 9. Crash cart with chest tube tray and Needle decompression kits or 14G needles (longer length) [stored outside room in the hallway until called for]. 10. Traumaman on wheeled cart with fresh skin in place 11. PPE equipment (i.e., gloves, sink or alcohol gel, face shield, disposable gowns) 12. IV start kit including tape (modified for us on an SP) 13. IV tubing 14. IV fluids (1 liter bag of NS) 15. Antibiotic piggyback with tubing (Vancomycin) For the simulated patient (Call 1 hour prior to scenario start): 16. Simulated patient wearing a patient gown. 17. Patient ID band 18. Moulage of right chest area - abrasions and redness where he struck the sink. 19. Moulage right chest redness, abrasions and evidence of cellulitis (redness) on left lower leg. 20. Kerlex dressing around L foot as evidence of drained abscess on dorsal aspect of the left foot (no moulage underneath needed). For the primary nurse:

	21. iAuris Stethoscope to support simulated differences in lung sounds. Will hand to participant during scenario. 22. Patient chart with appropriate admission documentation, labs, orders and nursing treatment. 23. 12 Lead ECG showing sinus tachycardia a. May be requested by participants to be ready at arrival to the patient room or made ready within 3-5 minutes after request by the SP portraying the tech. Other considerations: 24. Chest X Ray a. If requested, the primary nurse will leave to call and return and tell them they will be about 15-20 minutes -- they are currently in the SICU. 25. EFAST Ultrasound a. If requested, indicate to the participant that someone has gone to get the device. b. Ultrasound won't become available during the scenario. 26. Other Radiology Studies a. If requested, the primary nurse will ask them to put the order in CHCS -- "your going to have to put the order in CHCS" b. If the primary nurse needs to call down to radiology they can share that they've got a patient in the scanner right now or something to this effect. 27. Laboratory Studies a. If requested, the primary nurse or tech will simulate drawing blood and state they will send to the lab. b. When RRT team arrives they may bring an iStat with them; however, we will terminate the scenario after the participant gives the RRT nurse report. 28. ABG a. If requested, the tech can say they will call respiratory. b. If we make this available we will need to develop some pre-staged ABG readings at predetermined time markers (Time 0, 5 min, 10 min, 15 min,20 min)
What props would further support the clinical situation?	1. For needle decompression/chest tube insertion we will utilize a hybrid strategy using Trauma Man with skins for needle decompression or chest tube insertion.

What personal patient safety equipment should be available for the scenario?	1. Gloves 2. Alcohol gel 3. Gowns 4. Face shield
What are the diagnostic findings that would be needed to support participants as they make or confirm a diagnosis in this scenario?	1. Narrative and injury that is plausible enough to cause a tension pneumothorax (see storyboard) 2. Simulated patient portraying an increasingly anxious and uncomfortable patient using verbal and gestural cues. 3. Vital signs showing a trending decline in blood pressure, rising heart rate, declining oxygen saturation and diminished breath sounds on the affected side. (See storyboard for trends) 4. Chest X Ray images showing rib fracture and pneumothorax. 5. Admission documents and labs
What diagnostic activities (e.g. auscultation, palpation, etc.) would normally be used in this type of scenario?	1. Visualization 2. Auscultation 3. Palpation 4. Review and interpret consecutive vital signs readings. 5. Request and interpret chest X Ray 6. Request and interpret a 12 Lead 7. Diagnostic questioning 8. Situational management
What types of therapies (fluid challenge, medications, etc.) would typically be offered in this type of scenario?	In place prior to scenario start: 1. Peripheral IV a. IV start kit including tape (modified for us on an SP) b. IV tubing c. IV fluids (1 liter bag of NS) d. Antibiotic piggyback with tubing (Vancomycin) e. Patient ID band Available for use as scenario progresses? 1. Additional peripheral IV a. IV start kit b. IV tubing c. IV fluids (type of fluid prefered?) 2. Needles or needle decompression kits 3. Chest Tube Kits 4. Chest Tubes 5. Pain management medication options. 6. Pleur Evac (only need 1 - can be reused)

Roles and Rules

Question	Free Text Response

What rules would normally guide or govern care or behavior in this scenario?
 	
Rules –are conventions or guidelines that regulate activities. For example, is there a standard that governs how all patients who present with trauma are assessed or treated? Are there standard order sets that govern care?
	Assessment and Practice Guidelines:
	1. WRNNMC assumptions of practice for medicine and surgical,
	  1. Interns
	  2. Residents
	  3. Attendings
	2. ACLS Guidelines
	3. ATLS Guidelines
	4. WRNNMC guidelines for assigning patients to surgical and medical teams for inpatient coverage.
	5. WRNNMC RRT Guidelines
	6. Simulation scenario guidelines
	  1. Run in real time (except the time frame from call to arrival at patient door due to close proximity & potential need to send RRT team sooner).
	  2. Participants will be advised to perform the actions they feel are necessary. The SP nurse can remind them if they only use words.
	  3. We will ask participants to work within their own limitations and capabilities.
	  4. Rather than allowing the patient to fully decline, we will speed up the RRT team for those who are not comfortable with needle decompression or chest tube placement.

Who is typically present during a scenario such as this, and what role do they play during the event? Please give a brief description of what the role entails. Please include all roles typically present. Roles – are the division of labor. For example, so you may have a physician stationed at the head of the bed whose role is airway. The person present is the physician and their role is airway management.	1. Standard participant to portray the patient. 2. Standard participant to portray the patient's primary nurse. 1. Assist the participant in caring for the patient and to help with unfamiliar material or issues that may arise during the scenario so the scenario narrative can continue to advance.

	3. Standard participant to portray a backup\ nurse or technician role 1. Assist with information gathering outside the room, bringing the patient chart, requesting additional resources (e.g., radiology) and helping set up for a needle decompression/chest tube placement (which uses a hybrid simulation strategy). 4. One individual to portray the attending or senior resident for participants to call and request help from. 5. One individual to portray the RRT ICU nurse (RRT team member arrival is often staggered as people arrive from different locations in the hospital) member - first to arrive to receive handoff. (Cameo role).

Simulated Participant (Patient)

Domain: Inpatient ward

Diagnosis: Pneumothorax →Tension Pneumothorax

Name: Carl Fisher

Gender: Male

Age: 50

Unique Findings/Characteristics: None

Compatible Characteristics: N/A

Setting: inpatient ward (admitted to 4 Center around 0500 the morning of the scenario)

Preadmission Backstory:
Presented to the ER last night with foot pain, erythema, and swelling for the past 5 days. Also reported fevers for the past 3 days. He may have scraped or injured the area while working in the yard, but is not sure. He has been having a lot of pain in his foot especially while walking and wearing shoes, he has been limping and stayed home from work yesterday (the day of ED presentation). Reports fevers at home up to 100.6 for the past 3 days and some chills. Mostly healthy, has a history of diabetes and hypertension. Normally exercises regularly, but has not since this foot pain started.

In the ER, I&D (incision and drainage) of the abscess was performed late last night. The wound was packed with gauze and wrapped in a bandage with your sock placed over it. You were given Tylenol in the ER for pain and fever.

General Appearance: At the outset the patient is sitting on the bed, in his patient gown, not yet wearing specific fall prevention socks (having only just arrived on the floor. He initially appears as if he is protecting his chest where he hit the sink because movement, palpation and deep breaths make the pain much worse. He's also just generally frustrated and tired after being sick for the last 5 days, 3 of them with a fever that didn't resolve and eventually led him to seek care in the ED. He didn't expect to need to be admitted.

Clothing/Gown: in a patient gown with SP belongings in a patient belongings bag either sitting on the chair in the room or hanging on the backdoor.

Moulage:

- Evidence of redness on his right lower chest where he struck the sink– mostly red with some purple present – this may vary some depending on the SP's skin tone. Slight abrasions also present.
- Redness related to cellulitis in the lower left leg and dressing (Kerlex wrapped around foot and taped in place) over a drained abscess on dorsum of foot (near space between the toes) – no moulage under the wound dressing.

Chief Complaint: Shortness of breath, chest pain with inspiration and movement that is worsening following trip and fall when exiting the bathroom. No loss of consciousness reported and did not hit his head.

Opening Statement: See opening stem on page 1.

History of Present Illness: See opening stem on page 1.

Affect and Behavioral Expectations: He's generally frustrated and tired after being sick for the last 5 days, 3 of them with a fever that didn't resolve and eventually led him to seek care in the ED. He didn't expect to need to be admitted.

As scenario progresses [about 5 minutes after the physician enters the room] his fatigue and frustration begin to make way to his feeling increasingly concerned and anxiousness as the tension pneumothorax develops. Towards minute 12 Mrs. Fischer is really struggling to breath and feels like no matter how hard he tries he can't get a good breath and likens it to feeling like he's breathing through a straw.

In the final phase of the scenario (starting around minute 15) the patient will begin to become less anxious and is "running out of steam." He's hypoxic (low oxygen and rising C02) at this stage, may feel a little dizzy and his verbal responses are slowed and not clear to others in the room. Participants may do a number of things at this stage, including:

- Focus on stabilizing him with oxygen and start another IV while awaiting help from the rapid response team. In this case the study team will not allow the scenario to run long enough to where the patient could develop cardiac arrest.
- Perform a needle decompression where we will ask the physician participants to do the procedure on a model but encourage to interact with the SP patient (hybrid). Should participants do this you will feel better very quickly as you "catch your breath" though you still have a lot of pain where you struck the sink.
- Place a chest tube where we will ask the physician participants to do the procedure on a model but encourage to interact with the SP patient (hybrid). Should participants do this you will feel better very quickly as you "catch your breath" though you still have a lot of pain where you struck the sink.

Onset of complaint: Approximately 5 - 10 minutes prior to the nurse calling the physician to assess the patient.

Location of Pain: Pain is on the right side of the chest where the patient struck the sink.

Character/Quality of Pain/Complaint: Feels sharp, especially when breathing in. Cannot raise his hands above his head.

Severity/Pain Scale:

- 7/10 (currently) sharp pain when he tries to take a deep breath or moves *(gesture should include shorter breaths - inspirations)*
- If the patient is able to sit still and minimize movement the pain is an achy 3/10 but he fears moving (gestures will include guarding this area and wince with any effort to palpate.
- The pain decreases some to 5/10 with pain management medication if given IV - if given by mouth pain will remain unchanged because it won't have time to take effect during the scenario time.

Frequency of Pain: Constant aching pain since the fall that gets much worse with breathing or movement because he fracture a couple of ribs (6/7/8 - lower ribs) when he hit the sink.

Aggravating Factors: worsens significantly with inspiration and exaggerated arm movement or movement in general. When the SP moves he should do so in a "guarded" manner (gestural cue) – bracing or holding (gestural cue) his chest where he struck the sink. If the physician tries to examine his chest the SP should be guarded as they pull back the gown and guard the site if the physician tries to palpate the injured area. Laying down makes it harder to breath and as the tension pneumothorax gets worse the SP should be resistant to lying back.

Alleviating Factors: Initially felt better by bracing with a pillow or his arm. The SP should brace and be protective of the right side which will help make the ribs more stable and move less, thus lessening the pain.

Associated Symptoms: Shortness of breath – initially at the scenario outset that he's feels pain where he struck his chest on the sink - but this eventually worsens to his feeling like it's getting harder to breathe.

Similar Episodes: No

Admission Medications:

- Vancomycin 1500 mg q6hr - first dose given 30 min prior to scenario start

Pre-Admission Medications:

- HCTZ 25 mg every morning once per day)
- Prilosec 20 mg every morning once per day)
- Metformin 500 mg in the morning and evening twice daily

Past Medical History: Diabetes (5 years), HTN (8 years), GERD

Surgical History: No prior surgical history.

Social History (Shx):

Occupation: Chaplain
Religion: Episcopal
Education: B.A. History, M. Divinity
Length of Military Service: [Need]

Habits:

Alcohol: 2-4 drinks per week (combination of wine or Scotch on occasion), CAGE = 0/4
Tobacco: 1/2 pack per day x 5 years but quit 25 years ago
Drug use: Denies
Sexual History: Not currently sexually active

Family History:

Father died of PE age 55 after gallbladder surgery
Mother died of Breast cancer age 75
No siblings
Widowed
3 children (aged 24, 22, 18) alive, all well

Physical Exam Results:

- **Central nervous System**: Alert, oriented, denies loss of consciousness, denies striking his head, anxious at the beginning - becomes increasingly anxious and frustrated.
- **Cardiovascular:** Tachycardia, regular rhythm
- **Pulmonary**: Tachypneic, Breath sounds on right chest diminished
- **Gastrointestinal**: some general diffuse abdominal pain 3/10
- **Musculoskeletal**: Pain @ injury site of right lateral thorax
- **Integumentary**: Intact, redness (recent bruising minor abrasions) noted at injury site of right thorax and right elbow/hand from the fall. Also, redness related to cellulitis in the lower left leg and dressing over a drained abscess on dorsum of left foot (near space between the toes).
- **Psychological**: Anxious and agitated due to worsening difficulty breathing and hypoxemia.

Simulated Participant (Primary Nurse)

Domain:
Inpatient medical or surgical ward

Gender:
Male or female

Age:
Can vary

Goals of this Character:
The primary nurse role serves to support the scenario by performing requests from the physician participant, answer questions from the physician participant regarding the patient in the scenario, and to help prompt the participant in drawing out certain types of information to help the study team in gaining a better understanding of the study participant's perceptions while the scenario is still active. The primary nurse places the initial phone call to the physician participant, enters the patient room with the participant, and is present in the room throughout the scenario. The primary nurse is helpful and may perform some tasks, such as administering oxygen, without being asked. The primary nurse may also draw attention to information that may indicate the correct diagnosis, such as noting that he or she could not hear breath sounds on the right side in the event the iAurus stethoscope malfunctions or if the participant struggles for and extended period of time.

For the purpose of this scenario and the JPC simulation study, the lines are scripted for uniformity among participants, but may be adjusted to each situation. Ideally, this role is portrayed by a carefully trained standardized participant or a study team member for consistency.

General Appearance & Clothing:
Active duty nurses on med/surg floors normally wear their respective camouflage uniform (sometimes with a matching scrub top instead of the normal blouse). Civilian nurses wear either personal scrubs or hospital scrubs.

Goals of the Primary Nurse:
Give participant background information about the patient. Cue the participant as needed in assessing the situation, performing interventions, and calling for help.

Relevant Clinical Background:
Holds a BSN. Has worked as a Med-Surg nurse for 3 years.

Opening Statement:
"Hi, this is [name of nurse] up on 4 Center.

"I'm calling about Mr. Fisher – he was just admitted from the ED about 30 minutes ago with cellulitis. We got his antibiotics hung and he got up to go to the bathroom and fell. He says he got tangled up in his IV pole. He's alert and oriented, denies any loss of consciousness and says he didn't hit his head...but he's got a lot of pain on his chest where he hit the sink."

Other Potential Opening Scene Statements:
"I just helped him get back into bed but he's in a lot of pain."

"I haven't had a chance to get them (vitals)*...He just came up from the ED."*

"I was just getting ready to go in and do my assessment on him when he fell."

"*He was just admitted from the ED with cellulitis of the left lower leg and foot."*

"*I don't have one* (an EKG) *and...it looks like they didn't do one in the ED. Do you want one?"*

Rapid Response Team Nurse - 1st to arrive

Domain:
Inpatient medical or surgical ward

Gender:
Male or female

Age:
Can vary

Primary Purpose:
The rapid response team nurse role serves to support the scenario by engaging the participant about what has transpired leading up to their arrival. The rapid response nurse arrives as the scenario time is up, at approximately minute 13-15 and will seek to gather certain key information from the participant and then end the scenario. This participant is helpful.

For the purpose of this scenario and the JPC simulation study, the lines are carefully scripted to deliberately draw out certain types of information to help the study team in gain a better understanding of the study participant's perceptions while the scenario is still active. For example, the participant in this role will ask about what has transpired, what the participant thinks the patient's problem is, and what is their reasoning for treatment thus far. Ideally this role is portrayed by a carefully trained standardized participant or a study team member for consistency.

Clothing:
Scrubs or related hospital attire

Opening Statement:
"Hi, I'm (insert your name) from the ICU. Someone called a rapid response for this patient? What's going on?"

Other Key Statements or Actions:
"What have you done for him so far?"

"What do you think is going on with him?"

"What are you thinking in terms of next steps?"

"Okay, sounds good. I can call the ICU to see about getting a bed assignment"

To End the Scenario:

"Okay, thanks. We're going to end the scenario here."

Last Revised 08 May, 2018

Outpatient, New Onset Diabetes – No Contextual Factors

Study Team Version

Primary Diagnosis:	Contextual Factors:
New Onset Diabetes Mellitus Alternative Diagnosis: Diabetes Insipidus SIADH DKA Pituitary Adenoma Hyperthyroidism Urinary Tract Infection	No EHR *Complex Case Presentation

Setting: Outpatient Clinic

Name: Michelle DeMoro

Age: 45 Years

Date of Birth: October 11, 1971 (may vary based on actual casted characteristics)

Height/Weight: Use standardized patient's

SP Characteristics for Patient Desired: Non-hispanic female, slightly overweight

Summary Case Stem (see SP details and storyboard for in-depth information):
This scenario is set in the outpatient acute care clinic exam room and begins when the physician enters to meet Mrs. Demoro. Mrs. Demoro is a 45 year old dependent of an active duty Navy Officer. Mrs. Demoro is slightly nervous and embarrassed about her symptoms, including frequent urination and excessive thirst. She carries a water bottle with her and occasionally drinks from it during the interview.

Simulated Participant Details

Domain: Acute Care Clinic

Diagnosis: New Onset Diabetes Mellitus

Alternative Diagnosis Participants may Explore: Diabetes Insipidus, DKA, SIADH, Hyperthyroidism, Pituitary Adenoma, Urinary Tract Infection

Contextual Factors:

- No EHR

Gender: Female

Age: Range 40-50

Unique Findings/Characteristics: Mrs. DeMoro is a good medical historian and seeks regular preventive care such as getting her thyroid levels drawn, getting blood pressure checks, and getting her annual well-woman exams. She doesn't recall specific details about her thyroid or blood pressure, in part because her blood pressure has been well controlled and her thyroid levels have been normal. Mrs. DeMoro's spouse is an active duty Naval officer who works in the intel community. They have one daughter who is 16 years old who plays a lot of sports after school. Today is her first visit to this clinic because they have recently transferred to the area. This is why her records are not available.

Setting: Outpatient Clinic Exam Room

General Appearance: Tired, occasionally drinks from her water bottle during the interview.
Somewhat reluctant to share some of the details about frequent urination and recurring yeast infections because she finds it embarrassing.

Clothing/Gown: Street clothes

Moulage: None unless participant needs to look tired.

Chief Complaint: Fatigue, constantly thirsty, frequently hungry, frequent urination

Opening Statement (Close but not verbatim):

"Well… [pause] I have been very thirsty lately. Lately, I just can't seem to quench my thirst… [thinking] I have been carrying this bottle with me everywhere for the last 2 weeks, and probably fill it up 6 times a day (16 ounces) If that wasn't enough, I feel like I am tired all the time "

History of Present Illness: Mrs. Michelle DeMoro is a 45-year-old presents with fatigue, increased thirst and appetite, and frequent urination for the past several weeks. The fatigue and thirst have come on gradually and have not gone away or gotten better. She also complains of intermittent blurred vision but passes it off as needing to go have her vision checked for a new pair of glasses (she doesn't make the connection between blurred vision and diabetes).

Pertinent Medical History: Flu 6-8 weeks ago (self-diagnosis - she didn't get a formal flu test).

Affect and Behavioral Expectations: Cooperative, good medical historian, embarrassed about having frequent urination but is willing to discuss with the physician.

Location of Pain: No pain

Character/Quality of Pain/Complaint: N/A

Severity/Pain Scale: N/A

Frequency of Pain: N/A

Aggravating Factors: N/A

Alleviating Factors: None

Associated Symptoms: N/A

Similar Episodes: No

Current Medications:

- HCTZ 25 MG once daily in the morning
- Synthroid [Not sure of the dose - but dose hasn't changed in a long time if asked - has been taking this for about 2 years] in the morning
- Gyne Lotrimin Cream as needed for yeast infections - they've helped some.

Past Medical History:

- High blood pressure (hypertension) for several years
- Parathyroid surgery for Adenoma (2 years ago). These are not related but they did occur around the same time period.
- Goiter 2 years ago
- Automobile accident 4 years ago - hospitalized briefly (1-2 days) with a head injury. She was driving and the other driver blew the stop light and T-boned her. Mrs. Demoro struck her head on the driver's-side window,shattering it. She hasn't had any issues since she was discharged.
- Last well woman visit and mammogram were about 6-8 months ago - both were normal.
- Last menstrual period - about 2 weeks ago (or use your own dates)

Social History:

Education: Bachelor's degree
Occupation: Retail Sales - works part -time at the Container Store
Religion: Unknown
Lifestyle: Dependent of an active duty Navy Officer (Intel), has one daughter, aged 16 years. Feels like she is always on the go getting her daughter to various sports practices and games.

Habits:

Alcohol: Quit 2-3 years ago because she didn't really like the taste. Prior to this she would have 1-2 drinks (wine/beer) during social occasions 1-2 times per month.
Tobacco: Quit 2-3 years ago - 1 pack per day for 20 years [20 pack per year]
Drugs: Denies
Diet: Moderately high-fat, drinks 1 -2 cups of coffee/day, eats on the run a lot.
Exercise: Doesn't have a lot of free time.
Sexual History: Married, Sexually active – however has had three recent yeast infections which have made sex painful. Using Gyne Lotrimin with some relief.

Family History: Father died @ age 55 from a heart attack, mother has history of TENS (SP is uncertain about exactly what this diagnosis is).

Physical Exam Results:

- CNS: Alert, oriented
- HEENT: Normal, vision exam will be normal
- Cardiovascular: regular rhythm, no murmurs, rubs, gallops
- Pulmonary: Clear to auscultation
- Gastrointestinal: normal
- Musculoskeletal: Normal
- Integumentary: Normal
- Psychological: Normal

Vitals:

- Pulse: 94
- BP: 145/85 (this is what we will tell the study participant - this is a little higher than usual for you)
- Respirations: 16
- SP02: 99% RA
- Temp: 98.8
- Lungs – clear bilateral

Labs:

- Labs within 6 months:
  - Hemoglobin = 10 gm/dL
  - Basic Metabolic Panel = normal
  - A1C=6.5
  - HIV - Negative
  - Other records - Well woman - normal, immunizations current

SP Statements (do not have to adhere verbatim - but suggested)

Opening statement:
Mrs. DeMoro: Well… [pause] I have been very thirsty lately. Lately, I just can't seem to quench my thirst… [thinking] I have been carrying this bottle with me everywhere for the last 2 weeks, and probably fill it up 6 times a day (16 ounces) If that wasn't enough, I feel like I am tired all the time… [thinking]

Additional statements that may be used regarding symptoms:
Mrs. DeMoro: I think my thirst has been more gradual...and I first noticed feeling tired a few months ago when I had the flu.

Mrs. DeMoro: [pause, embarrassed] I feel like all I do is go to the bathroom all day long… and not just little amounts… [emphatic] I mean 7 or 8 bladders full every day and at least 3 or 4 times at night… [looking down, upset] I feel like a human water filter.

Appetite:
Mrs. DeMoro: I have been hungrier lately and I've been eating a lot more than usual but haven't gained any weight.

If asked about vision:
Mrs. DeMoro: [thinking] This is a little off the wall, but I have also noticed that my vision seems to be blurry sometimes… [matter of fact] probably just time to see the optometrist again for a new prescription.

Prior Surgeries:
Mrs. DeMoro: I had parathyroid surgery about 2 years ago for an adenoma. I was also in a serious car crash about four years ago and was hospitalized with a concussion….but it hasn't given me problems for years.

Family medical history:
Mrs. DeMoro: My father passed when he was about 55 from a heart attack. My mother has a history of TENS—something about endocrine problems.

Additional Concerns - Yeast Infection:
Mrs. DeMoro: This is a little embarrassing, but it is starting to affect my marriage…[pause]…I've developed three yeast infections over the past couple of months,and love-making has become very painful. I used that lotra-something medication which seems to help a little.

Proposed Storyboard	
0-3 min **Hand -Off and Scenario Start** 	1. After the participant receives their instructions and indicates they have no further questions (see participant hand off script below) they will receive details about their patient in the form of doorway information with patient details, i.e., reason for today's visit, vitals, past records (6 mos). ***This will be given to them by a study coordinator using a standardized document.** 2. The participant will have time to review this document and then will determine when they are ready to enter the clinic exam room where the SP will be waiting (in street clothes). This is the start of the live scenario. 3. Note - the scenario will run in real time rather than speeding the time lapses up. We reason that by running the scenario in actual time will more accurately allow us to assess the participant's clinical reasoning processes. Speeding time up sets up a situation in which we could skew participant's reasoning by making time move more quickly than typically experienced in the actual clinical setting. 4. Note - participants will only have access to those diagnostic findings that they seek and perform. Diagnostic findings will not be volunteered without prompting because this will allow us to assess their reasoning processes, including, what they prioritize early in their assessment, what diagnostic evidence they use to make choices, and the order in which they ask for and use it.

Time	Manikin Settings and Changes	Patient Responses/Cues	Potential Participant Actions
1-10 min **Initial Interview** 	**Vitals:** Pulse: 94 BP: 145/85 Respirations: 16 SP02: 99% RA Temp: 98.8 Lungs – clear bilateral	**Affect:** Appears concerned and somewhat anxious. **SP Opening Statement:** Follows after the SP and physician exchange greetings and the physician asks how he is doing or what brought him to the clinic today. SP:"Well… [pause] I have been very thirsty lately. Lately, I just can't seem to quench my thirst… [thinking] I have been carrying this bottle with me everywhere for the last 2 weeks, and probably fill it up 6 times a day (16 ounces) If that wasn't enough, I feel like I am tired all the time "	1. Participants should conduct an initial interview. 2. The interview should give way to a physical exam.
5-15 min **Physical Exam** 	**Vitals:** Pulse: 94 BP: 145/85 Respirations: 16 SP02: 99% RA Temp: 98.8 Lungs – clear bilateral	**Affect:** Appears somewhat embarrassed but embarrassment is within her - not as a function of her relationship with the physician. If asked about other symptoms or anything not discussed: "This is a little embarrassing, but it is starting to affect my marriage…[pause]…I've had three	1. Reviews/analyzes vital signs 2. Continues communication with the patient (diagnostic questions, education and counseling) 3. May use a snellen chart or other eye chart to screen vision.

		yeast infections over the past couple of months, and love-making has become very painful."	
15 min **Scenario End -** 	Scenario End	Scenario End	1. Facilitator should end the scenario if the participant hasn't done so yet by entering the room and announcing that we will stop here

Scenario Context Questionnaire

Question	Free Text Response
What are the typical physical tools (e.g., stethoscope, ultrasound device, etc.) that are needed or found in this type of scenario?	1. Outpatient exam bed - sheets or paper 2. Patient Gown(available but not worn) 3. Stethoscope 4. Ophthalmoscope (on headwall) 5. Otoscope (on headwall) 6. Otoscope covers 7. Blood Pressure Cuff 8. Pocket Snellen Chart 9. Cotton tipped applicators 10. Clock on wall
What props would further support the clinical situation (e.g., standard patient with moulage on the [location], human patient simulator, etc.)?	1. Water bottle 2. Standard Patient
What personal patient safety equipment should be available for the scenario?	1. Working sink 2. Soap 3. Paper towels 4. Gloves 5. Alcohol gel for hands 6. Alcohol prep pads 7. Alcohol wipes (for cleaning between scenarios)
What are the diagnostic findings that would be needed to support participants as they make or confirm a diagnosis in this scenario?	1. Vital signs
What diagnostic activities (e.g. auscultation, palpation, etc.) would normally be used in this type of scenario?	1. Visualization 2. Auscultation 3. Palpation 4. Obtain blood pressure 5. Diagnostic questioning
What types of therapies (fluid challenge, medications, etc.) would typically be offered in this type of scenario?	1. Social and emotional support 2. Education and counseling

Roles and Rules

Question	Free Text Response

What rules would normally guide or govern care or behavior in this scenario? Rules –are conventions or guidelines that regulate activities. For example, is there a standard that governs how all patients who present with trauma are assessed or treated? Are there standard order sets that govern care?	Assessment Guidelines: 1. Limited physical exam Therapeutic Guidelines: 1. Social and emotional support 2. Education and counseling
Who is typically present during a scenario such as this, and what role do they play during the event? Please give a brief description of what the role entails. Please include all roles typically present. Roles – are the division of labor. For example, So you may have a physician stationed at the head of the bed whose role is airway. The person present is the physician and their role is airway management.	1. 1 Standard Patient

Author(s):
Steven Durning, MD, PhD, FACP
[Add Others]

Adaptation Authors:
Alexis Battista, PhD
Steven Durning, MD, PhD, FACP
Sarah Krajnik, RN, BSN
Anna Howle, MAC
Jeff Mikita, MD
Jeff La-Rochelle, MD
Paul Hemmer, MD, MPH

Live Scenario – Inpatient Trauma[1]

Summary:
This scenario is set in the inpatient setting where Mr. Carl Fisher, a 50-year-old male, was admitted the previous evening (around 0500) with a diagnosis of cellulitis. Today (time of the study) Mr. Fisher has tripped and fallen on the sink (right chest) in his bathroom which causes a rib fracture and subsequent pneumothorax. The fall was not witnessed, but he told the nurse that he got tangled up in the wires and feet of his IV pole and fell. After his fall, the nurse responded, helped him settle back to bed, asked the technician to check his vitals and called the physician to come check him out. At the beginning Mr. Fischer isn't in significant distress; however, as the scenario progresses the pneumothorax will progress to a tension pneumothorax that requires treatment with needle decompression or placement of a chest tube. Study participants are also expected to eventually call for a rapid response.

Leading Diagnosis	Differential Diagnoses	Contextual Factors
❖ Pneumothorax → Tension Pneumothorax	❖ Rib Fracture ❖ Hemothorax ❖ Pulmonary Contusion ❖ Pulmonary Embolism ❖ Syncope related to cardiac arrhythmia (e.g., a fib, Vtach) ❖ Hypovolemia (e.g., dehydration, sepsis) ❖ Anaphylaxis ❖ Stroke/CVA ❖ Pneumothorax ❖ Cardiac Contusion ❖ Splenic or liver laceration/hematoma	❖ *Limited knowledge of the patient* ❖ *Emotional volatility due to increasing hypoxia* ❖ *Increasing acuity of presentation* ❖ *Team-based clinical reasoning*

[1] Author(s):
Alexis Battista, PhD
Steven Durning, MD, PhD
Anna Howle, MAC
Sarah Krajnik, RN, BSN
Jeff La-Rochelle, MD, MPH
Paul Hemmer, MD, MPH
Jeff Mikita, MD
Walter Kucera, MD
Matthew Nealeigh, DO
Sarah Ordway, MD
Thomas Mellor, MD
Megan Ohmer, BS

Last Revised 08 May, 2018

Scenario Storyboard			
Scenario Start  **0-3 min**	1. After the participant receives their instructions and expectations for participants and indicates they have no further questions (see participant hand off script below) they will receive a phone call (using a study team designated phone) from the participant portraying the patient's primary nurse. 2. **Opening Statement** [Ideally read verbatim for continuity by SP Primary Nurse]: *"Hi, this is [name of nurse] up on 4 Center. I'm calling about Mr. Fisher – he was just admitted from the ED about 30 minutes ago with cellulitis. We got his antibiotics hung and he got up to go to the bathroom and fell. He says he got tangled up in his IV pole. He's alert and oriented, denies any loss of consciousness and says he didn't hit his head...but he's got some pain on his chest where he hit the sink."* His vital signs are: *Pulse: 98* *BP: 130/80* *Respirations: 18* *SP02: 98% RA* *Pain: 7/10 at the injury site* 3. The study participant can ask further questions of the primary nurse, such as how is he doing now? What are his vital signs? Can you get a 12 lead so I can see it when I get there? If the participant asks for a 12 Lead let the participant know you will work on getting one. See details for Primary Nurse for additional responses. 4. Following the call - the study coordinator will direct the participant to the patient's inpatient room.[2]		
Time	**Goals and monitor settings**	**Patient Responses/Cues**	**Potential Participant Actions**

Commented [1]: +anthony.artino@usuhs.edu Tony, This is the scenario storyboard for the new inpatient trauma scenario for you to review so you can get a better understanding of what participants will be expected to do as this scenario unfolds. I will share the specific questions that we've designed so far under separate cover. I also sent a link to Tim.
Assigned to Anthony Artino

Commented [2]: +tc454@rci.rutgers.edu Tim, This is the scenario storyboard for the new inpatient trauma scenario for you to review so you can get a better understanding of what participants will be expected to do as this scenario unfolds. I will share the specific questions that we've designed so far under separate cover.
Assigned to tc454

Commented [3]: This could be a good opportunity to study their forethought processes and should be recorded and re-played as a part of the think aloud.

[2] The scenario will run in real time rather than speeding the time lapses up. We reason that by running the scenario in actual time will more accurately allow us to assess the participant's clinical reasoning processes. Speeding time up sets up a situation in which we could skew participant's reasoning by making time move more quickly than typically experienced in the actual clinical setting. Participants will only have access to those diagnostic findings that they seek and perform. Diagnostic findings will not be volunteered without prompting because this will allow us to assess their reasoning processes, including, what they prioritize early in their assessment, what diagnostic evidence they use to make choices, and the order in which they ask for and use it.

Physician arrival at bedside.  **3-8 min**	Goal: During this segment the patient will compensate to give the participant an opportunity to assess the patient and consider potential differential diagnoses. Vital Signs: *Pulse: 106* *BP: 124/78* *Respirations: 20* *SP02: 97% RA* **For the primary nurse SP – for vital signs checks in this stage of the scenario please take the cuff and SPO2 probe off the patient.	1. Patient is in the patient bed with staff having already obtained a set of vital signs at the participants' arrival. 2. In pain, cooperative, appears uncomfortable – braces *(holds area with arm or hand type gesture*) his injured side and tries to minimize movement (*stiff - guarded movement*), respiratory effort is mildly elevated and not deep because it's painful to take in a full breath. - Pain without movement is achy and 4/10 -- but "*fears*" the sharper pain that comes with movement. - Pain worsens - sharp - with movement or deep breath (8/10). - Increased pain and tenderness to the right lateral chest with palpation (*SP will guard if palpation attempted*) (8/10).	1. Communication with the patient (introduction, diagnostic questions). 2. Conduct an initial physical exam (focused). 3. Reviews/analyzes initial set of vital signs. 4. May request repeat vital signs. 5. Request and review patient admission documents, diagnoses, prior dx, and medications. 6. May request supplemental Oxygen, 7. May request pain management (e.g., Tylenol, Tylenol with a narcotic PO or medication by IV). 8. May request an ultrasound machine.
Time	**Goals and monitor settings**	**SP Responses/Cues**	**Potential Participant Actions**
Continued assessment,	Goal:	1. Pain may be decreased (5 or 6/10) if analgesia	1. Reviews/analyzes vital signs

Commented [4]: This segment will allow us to examine how they gather information from the patient (verbally and gesturally), the primary nurse and the medical record-- and the PEF/TA will help contextualize their reasoning.

Commented [5]: This scenario will also allow us to study how their differential diagnoses evolve from call to initial assessment and again during increased acuity.

patient deteriorates and rising acuity  **8 - 18 min**	During this time frame the patient will start to deteriorate which will further introduce increased acuity and increase patient anxiety. Vitals (trending down to the following): Pulse: 130 BP: 86/60 Respirations: 26 SP02: 91% with supplemental Oxygen. *Decreased breath sounds on the right (injury site). * *For the primary nurse SP – for vital signs checks in this stage of the scenario please leave the cuff and SPO2 probe on the patient.	previously given (e.g., Morphine). 2. Difficulty breathing continues to worsen even with supplemental oxygen. 3. Anxiety continues to the point where the patient occasionally is frustrated with the medical providers, including the study participant. 4. If the participant doesn't recognize the changes in patient affect and vital signs, the SP can elevate their level of frustration with the study participant. 5. Near the end of this stage the patient will start to become somnolent. 6. At the point of somnolence and if no RRT team called yet the SP nurse will cue the participant to call an RRT.	2. Continues communication with the patient. 3. May request supplemental Oxygen (e.g., nasal cannula, ox mask, simple mask, NRB). 4. May request subsequent vital signs checks (at this stage of the scenario the primary nurse will leave the BP cuff on) 5. May request moving crash cart into the room along with continuous 3 - lead ECG monitoring with a monitor 6. May request a stat chest x ray 7. May request an ultrasound machine. 8. May request additional vascular access be started 9. May choose to call rapid response or senior resident. 10. May decide to treat the tension pneumothorax and perform a need decompression or a chest tube.
Time	**Goals and monitor settings**	**Patient Responses/Cues**	**Potential Participant Actions**
Option 1: Participant chooses to support &	Goal: During this time frame the first of the rapid response teams will arrive	1. In the event the participant does not know how or feel comfortable performing a needle	1. The participant will give handoff report to the arriving RRT team member.

Commented [8]: This would be a good opportunity to insert a series of microanalysis questions that the primary nurse or the ICU nurse asks, such as "what are you thinking is going on?"

await RRT Arrival  **XX - 20 min**	and seek an initial report on the patient's status. The goal will be to allow this reporting to be completed because it represents an opportunity to learn about their understanding of the situation. After report the study team will stop the scenario. Vitals (trending depends on actions): Pulse: 136 BP: 84/60 Respirations: 26 SP02: 90% with supplemental Oxygen.	decompression or placement of a chest tube (even after cueing) the patient's condition will continue to deteriorate. 2. The RRT team will focus on getting report.	
Time	**Goals and monitor settings**	**Patient Responses/Cues**	**Potential Participant Actions**
Option 2 & 3: Needle decompress/Chest tube & await RRT Arrival to Handoff to	Goal: During this time frame if the participant chooses to treat the tension pneumothorax the	1. If they request a chest tube tray or a needle for decompression the primary nurse in the scenario will respond:	1. The participant needle decompresses or places a chest tube. 2. The participant may request a RRT – the primary nurse will cue this if not called and

Commented [6]: This will be the equivalent to an unprompted think aloud when they give hand off – similar to their talking with the patient about possible diagnosis in the other scenarios; however, here they will be talking with a fellow healthcare professional so that might be different.

Commented [7]: It might also give us a better idea of cognitive load because, in speaking with another health care professional, you'd expect the participant to use advanced "semantic qualifiers" and if s/he doesn't, that could indicate higher load.

End the Scenario  **XX – 20ish min**	patient will begin to improve. Shortly thereafter, the first of the rapid response team members will arrive and seek an initial report. Following report, the study team will end the scenario. Vitals (trending upwards): *Pulse: 120* *BP: 94/70* *Respirations: 22* *SP02: 94% with supplemental Oxygen.*	*"I'm going to call an RRT"* (if they don't call one) *"Let me get the crash cart"* (if they call for a chest tube/needle) 2. If the participant decompresses or places a chest tube the patient will experience pain during the procedure but will experience relief after placement.	request to do a decompression/chest tube. 3. May request regular vital signs assessments/updates. 4. The participant will give handoff report to the arriving RRT team member to allow for an opportunity to learn about the participant's understanding of the situation.

Simulated Activity System Questionnaire

Tools and Structured Interventions

Question	Free Text Response
What are the typical physical tools (e.g., stethoscope, ultrasound device, etc.) that are needed or found in this type of scenario?	For the patient room: 1. Inpatient med/surg room (med/surg ward) 2. Inpatient bed (med/surg ward) 3. IV pole x 1 (the one he tripped on) 4. Bedside table 5. Headwall with air/oxygen ports -simulated 6. Clock 7. Fresh linens and 2 gowns for SP 8. BP Cuff/Sp02 probe with automated device that can assess pulse, pulse ox, blood pressure with pre-printed & laminated vital signs ready/stored for easy access. 9. Crash cart with chest tube tray and Needle decompression kits or 14G needles (longer length) [stored outside room in the hallway until called for]. 10. Traumaman on wheeled cart with fresh skin in place 11. PPE equipment (i.e., gloves, sink or alcohol gel, face shield, disposable gowns) 12. IV start kit including tape (modified for us on an SP) 13. IV tubing 14. IV fluids (1 liter bag of NS) 15. Antibiotic piggyback with tubing (Vancomycin) For the simulated patient (Call 1 hour prior to scenario start): 16. Simulated patient wearing a patient gown. 17. Patient ID band 18. Moulage of right chest area - abrasions and redness where he struck the sink. 19. Moulage right chest redness, abrasions and evidence of cellulitis (redness) on left lower leg. 20. Kerlex dressing around L foot as evidence of drained abscess on dorsal aspect of the left foot (no moulage underneath needed). For the primary nurse:

	21. iAuris Stethoscope to support simulated differences in lung sounds. Will hand to participant during scenario. 22. Patient chart with appropriate admission documentation, labs, orders and nursing treatment. 23. 12 Lead ECG showing sinus tachycardia a. May be requested by participants to be ready at arrival to the patient room or made ready within 3-5 minutes after request by the SP portraying the tech. Other considerations: 24. Chest X Ray a. If requested, the primary nurse will leave to call and return and tell them they will be about 15-20 minutes -- they are currently in the SICU. 25. EFAST Ultrasound a. If requested, indicate to the participant that someone has gone to get the device. b. Ultrasound won't become available during the scenario. 26. Other Radiology Studies a. If requested, the primary nurse will ask them to put the order in CHCS -- "your going to have to put the order in CHCS" b. If the primary nurse needs to call down to radiology they can share that they've got a patient in the scanner right now or something to this effect. 27. Laboratory Studies a. If requested, the primary nurse or tech will simulate drawing blood and state they will send to the lab. b. When RRT team arrives they may bring an iStat with them; however, we will terminate the scenario after the participant gives the RRT nurse report. 28. ABG a. If requested, the tech can say they will call respiratory. b. If we make this available we will need to develop some pre-staged ABG readings at predetermined time markers (Time 0, 5 min, 10 min, 15 min,20 min)
What props would further support the clinical situation?	1. For needle decompression/chest tube insertion we will utilize a hybrid strategy using Trauma Man with skins for needle decompression or chest tube insertion.

What personal patient safety equipment should be available for the scenario?	1. Gloves 2. Alcohol gel 3. Gowns 4. Face shield
What are the diagnostic findings that would be needed to support participants as they make or confirm a diagnosis in this scenario?	1. Narrative and injury that is plausible enough to cause a tension pneumothorax (see storyboard) 2. Simulated patient portraying an increasingly anxious and uncomfortable patient using verbal and gestural cues. 3. Vital signs showing a trending decline in blood pressure, rising heart rate, declining oxygen saturation and diminished breath sounds on the affected side. (See storyboard for trends) 4. Chest X Ray images showing rib fracture and pneumothorax. 5. Admission documents and labs
What diagnostic activities (e.g. auscultation, palpation, etc.) would normally be used in this type of scenario?	1. Visualization 2. Auscultation 3. Palpation 4. Review and interpret consecutive vital signs readings. 5. Request and interpret chest X Ray 6. Request and interpret a 12 Lead 7. Diagnostic questioning 8. Situational management
What types of therapies (fluid challenge, medications, etc.) would typically be offered in this type of scenario?	In place prior to scenario start: 1. Peripheral IV a. IV start kit including tape (modified for us on an SP) b. IV tubing c. IV fluids (1 liter bag of NS) d. Antibiotic piggyback with tubing (Vancomycin) e. Patient ID band Available for use as scenario progresses? 1. Additional peripheral IV a. IV start kit b. IV tubing c. IV fluids (type of fluid prefered?) 2. Needles or needle decompression kits 3. Chest Tube Kits 4. Chest Tubes 5. Pain management medication options. 6. Pleur Evac (only need 1 - can be reused)

Roles and Rules

Question	Free Text Response

What rules would normally guide or govern care or behavior in this scenario?
 	
Rules –are conventions or guidelines that regulate activities. For example, is there a standard that governs how all patients who present with trauma are assessed or treated? Are there standard order sets that govern care?
	Assessment and Practice Guidelines:
	1. WRNNMC assumptions of practice for medicine and surgical,
	  1. Interns
	  2. Residents
	  3. Attendings
	2. ACLS Guidelines
	3. ATLS Guidelines
	4. WRNNMC guidelines for assigning patients to surgical and medical teams for inpatient coverage.
	5. WRNNMC RRT Guidelines
	6. Simulation scenario guidelines
	  1. Run in real time (except the time frame from call to arrival at patient door due to close proximity & potential need to send RRT team sooner).
	  2. Participants will be advised to perform the actions they feel are necessary. The SP nurse can remind them if they only use words.
	  3. We will ask participants to work within their own limitations and capabilities.
	  4. Rather than allowing the patient to fully decline, we will speed up the RRT team for those who are not comfortable with needle decompression or chest tube placement.

Who is typically present during a scenario such as this, and what role do they play during the event? Please give a brief description of what the role entails. Please include all roles typically present. Roles – are the division of labor. For example, so you may have a physician stationed at the head of the bed whose role is airway. The person present is the physician and their role is airway management.	1. Standard participant to portray the patient. 2. Standard participant to portray the patient's primary nurse. 1. Assist the participant in caring for the patient and to help with unfamiliar material or issues that may arise during the scenario so the scenario narrative can continue to advance.

	3. Standard participant to portray a backup\ nurse or technician role 1. Assist with information gathering outside the room, bringing the patient chart, requesting additional resources (e.g., radiology) and helping set up for a needle decompression/chest tube placement (which uses a hybrid simulation strategy). 4. One individual to portray the attending or senior resident for participants to call and request help from. 5. One individual to portray the RRT ICU nurse (RRT team member arrival is often staggered as people arrive from different locations in the hospital) member - first to arrive to receive handoff. (Cameo role).

Simulated Participant (Patient)

Domain: Inpatient ward

Diagnosis: Pneumothorax →Tension Pneumothorax

Name: Carl Fisher

Gender: Male

Age: 50

Unique Findings/Characteristics: None

Compatible Characteristics: N/A

Setting: inpatient ward (admitted to 4 Center around 0500 the morning of the scenario)

Preadmission Backstory:
Presented to the ER last night with foot pain, erythema, and swelling for the past 5 days. Also reported fevers for the past 3 days. He may have scraped or injured the area while working in the yard, but is not sure. He has been having a lot of pain in his foot especially while walking and wearing shoes, he has been limping and stayed home from work yesterday (the day of ED presentation). Reports fevers at home up to 100.6 for the past 3 days and some chills. Mostly healthy, has a history of diabetes and hypertension. Normally exercises regularly, but has not since this foot pain started.

In the ER, I&D (incision and drainage) of the abscess was performed late last night. The wound was packed with gauze and wrapped in a bandage with your sock placed over it. You were given Tylenol in the ER for pain and fever.

General Appearance: At the outset the patient is sitting on the bed, in his patient gown, not yet wearing specific fall prevention socks (having only just arrived on the floor. He initially appears as if he is protecting his chest where he hit the sink because movement, palpation and deep breaths make the pain much worse. He's also just generally frustrated and tired after being sick for the last 5 days, 3 of them with a fever that didn't resolve and eventually led him to seek care in the ED. He didn't expect to need to be admitted.

Clothing/Gown: in a patient gown with SP belongings in a patient belongings bag either sitting on the chair in the room or hanging on the backdoor.

Moulage:

- Evidence of redness on his right lower chest where he struck the sink– mostly red with some purple present – this may vary some depending on the SP's skin tone. Slight abrasions also present.
- Redness related to cellulitis in the lower left leg and dressing (Kerlex wrapped around foot and taped in place) over a drained abscess on dorsum of foot (near space between the toes) – no moulage under the wound dressing.

Chief Complaint: Shortness of breath, chest pain with inspiration and movement that is worsening following trip and fall when exiting the bathroom. No loss of consciousness reported and did not hit his head.

Opening Statement: See opening stem on page 1.

History of Present Illness: See opening stem on page 1.

Affect and Behavioral Expectations: He's generally frustrated and tired after being sick for the last 5 days, 3 of them with a fever that didn't resolve and eventually led him to seek care in the ED. He didn't expect to need to be admitted.

As scenario progresses [about 5 minutes after the physician enters the room] his fatigue and frustration begin to make way to his feeling increasingly concerned and anxiousness as the tension pneumothorax develops. Towards minute 12 Mrs. Fischer is really struggling to breath and feels like no matter how hard he tries he can't get a good breath and likens it to feeling like he's breathing through a straw.

In the final phase of the scenario (starting around minute 15) the patient will begin to become less anxious and is "running out of steam." He's hypoxic (low oxygen and rising C02) at this stage, may feel a little dizzy and his verbal responses are slowed and not clear to others in the room. Participants may do a number of things at this stage, including:

- Focus on stabilizing him with oxygen and start another IV while awaiting help from the rapid response team. In this case the study team will not allow the scenario to run long enough to where the patient could develop cardiac arrest.
- Perform a needle decompression where we will ask the physician participants to do the procedure on a model but encourage to interact with the SP patient (hybrid). Should participants do this you will feel better very quickly as you "catch your breath" though you still have a lot of pain where you struck the sink.
- Place a chest tube where we will ask the physician participants to do the procedure on a model but encourage to interact with the SP patient (hybrid). Should participants do this you will feel better very quickly as you "catch your breath" though you still have a lot of pain where you struck the sink.

Onset of complaint: Approximately 5 - 10 minutes prior to the nurse calling the physician to assess the patient.

Location of Pain: Pain is on the right side of the chest where the patient struck the sink.

Character/Quality of Pain/Complaint: Feels sharp, especially when breathing in. Cannot raise his hands above his head.

Severity/Pain Scale:

- 7/10 (currently) sharp pain when he tries to take a deep breath or moves *(gesture should include shorter breaths - inspirations)*
- If the patient is able to sit still and minimize movement the pain is an achy 3/10 but he fears moving (gestures will include guarding this area and wince with any effort to palpate.
- The pain decreases some to 5/10 with pain management medication if given IV - if given by mouth pain will remain unchanged because it won't have time to take effect during the scenario time.

Frequency of Pain: Constant aching pain since the fall that gets much worse with breathing or movement because he fracture a couple of ribs (6/7/8 - lower ribs) when he hit the sink.

Aggravating Factors: worsens significantly with inspiration and exaggerated arm movement or movement in general. When the SP moves he should do so in a "guarded" manner (gestural cue) – bracing or holding (gestural cue) his chest where he struck the sink. If the physician tries to examine his chest the SP should be guarded as they pull back the gown and guard the site if the physician tries to palpate the injured area. Laying down makes it harder to breath and as the tension pneumothorax gets worse the SP should be resistant to lying back.

Alleviating Factors: Initially felt better by bracing with a pillow or his arm. The SP should brace and be protective of the right side which will help make the ribs more stable and move less, thus lessening the pain.

Associated Symptoms: Shortness of breath – initially at the scenario outset that he's feels pain where he struck his chest on the sink - but this eventually worsens to his feeling like it's getting harder to breathe.

Similar Episodes: No

Admission Medications:

- Vancomycin 1500 mg q6hr - first dose given 30 min prior to scenario start

Pre-Admission Medications:

- HCTZ 25 mg every morning once per day)
- Prilosec 20 mg every morning once per day)
- Metformin 500 mg in the morning and evening twice daily

Past Medical History: Diabetes (5 years), HTN (8 years), GERD

Surgical History: No prior surgical history.

Social History (Shx):

Occupation: Chaplain
Religion: Episcopal
Education: B.A. History, M. Divinity
Length of Military Service: [Need]

Habits:

Alcohol: 2-4 drinks per week (combination of wine or Scotch on occasion), CAGE = 0/4
Tobacco: 1/2 pack per day x 5 years but quit 25 years ago
Drug use: Denies
Sexual History: Not currently sexually active

Family History:

Father died of PE age 55 after gallbladder surgery
Mother died of Breast cancer age 75
No siblings
Widowed
3 children (aged 24, 22, 18) alive, all well

Physical Exam Results:

- **Central nervous System**: Alert, oriented, denies loss of consciousness, denies striking his head, anxious at the beginning - becomes increasingly anxious and frustrated.
- **Cardiovascular:** Tachycardia, regular rhythm
- **Pulmonary**: Tachypneic, Breath sounds on right chest diminished
- **Gastrointestinal**: some general diffuse abdominal pain 3/10
- **Musculoskeletal**: Pain @ injury site of right lateral thorax
- **Integumentary**: Intact, redness (recent bruising minor abrasions) noted at injury site of right thorax and right elbow/hand from the fall. Also, redness related to cellulitis in the lower left leg and dressing over a drained abscess on dorsum of left foot (near space between the toes).
- **Psychological**: Anxious and agitated due to worsening difficulty breathing and hypoxemia.

Simulated Participant (Primary Nurse)

Domain:
Inpatient medical or surgical ward

Gender:
Male or female

Age:
Can vary

Goals of this Character:
The primary nurse role serves to support the scenario by performing requests from the physician participant, answer questions from the physician participant regarding the patient in the scenario, and to help prompt the participant in drawing out certain types of information to help the study team in gaining a better understanding of the study participant's perceptions while the scenario is still active. The primary nurse places the initial phone call to the physician participant, enters the patient room with the participant, and is present in the room throughout the scenario. The primary nurse is helpful and may perform some tasks, such as administering oxygen, without being asked. The primary nurse may also draw attention to information that may indicate the correct diagnosis, such as noting that he or she could not hear breath sounds on the right side in the event the iAurus stethoscope malfunctions or if the participant struggles for and extended period of time.

For the purpose of this scenario and the JPC simulation study, the lines are scripted for uniformity among participants, but may be adjusted to each situation. Ideally, this role is portrayed by a carefully trained standardized participant or a study team member for consistency.

General Appearance & Clothing:
Active duty nurses on med/surg floors normally wear their respective camouflage uniform (sometimes with a matching scrub top instead of the normal blouse). Civilian nurses wear either personal scrubs or hospital scrubs.

Goals of the Primary Nurse:
Give participant background information about the patient. Cue the participant as needed in assessing the situation, performing interventions, and calling for help.

Relevant Clinical Background:
Holds a BSN. Has worked as a Med-Surg nurse for 3 years.

Opening Statement:
"Hi, this is [name of nurse] up on 4 Center.

"I'm calling about Mr. Fisher – he was just admitted from the ED about 30 minutes ago with cellulitis. We got his antibiotics hung and he got up to go to the bathroom and fell. He says he got tangled up in his IV pole. He's alert and oriented, denies any loss of consciousness and says he didn't hit his head...but he's got a lot of pain on his chest where he hit the sink."

Other Potential Opening Scene Statements:
"I just helped him get back into bed but he's in a lot of pain."

"I haven't had a chance to get them (vitals)*...He just came up from the ED."*

"I was just getting ready to go in and do my assessment on him when he fell."

"*He was just admitted from the ED with cellulitis of the left lower leg and foot."*

"*I don't have one* (an EKG) *and...it looks like they didn't do one in the ED. Do you want one?"*

Rapid Response Team Nurse - 1st to arrive

Domain:
Inpatient medical or surgical ward

Gender:
Male or female

Age:
Can vary

Primary Purpose:
The rapid response team nurse role serves to support the scenario by engaging the participant about what has transpired leading up to their arrival. The rapid response nurse arrives as the scenario time is up, at approximately minute 13-15 and will seek to gather certain key information from the participant and then end the scenario. This participant is helpful.

For the purpose of this scenario and the JPC simulation study, the lines are carefully scripted to deliberately draw out certain types of information to help the study team in gain a better understanding of the study participant's perceptions while the scenario is still active. For example, the participant in this role will ask about what has transpired, what the participant thinks the patient's problem is, and what is their reasoning for treatment thus far. Ideally this role is portrayed by a carefully trained standardized participant or a study team member for consistency.

Clothing:
Scrubs or related hospital attire

Opening Statement:
"Hi, I'm (insert your name) from the ICU. Someone called a rapid response for this patient? What's going on?"

Other Key Statements or Actions:
"What have you done for him so far?"

"What do you think is going on with him?"

"What are you thinking in terms of next steps?"

"Okay, sounds good. I can call the ICU to see about getting a bed assignment"

To End the Scenario:

"Okay, thanks. We're going to end the scenario here."

Last Revised 08 May, 2018

Viewpoint

FREE

June 12, 2018

Management Reasoning
Beyond the Diagnosis

David A. Cook, MD, MHPE[1,2]; Jonathan Sherbino, MD, MEd[3]; Steven J. Durning, MD, PhD[4]

» Author Affiliations | Article Information

JAMA. 2018;319(22):2267-2268. doi:10.1001/jama.2018.4385

Clinical reasoning—the integration of clinical information, medical knowledge, and contextual (situational) factors to make decisions about patient care—is fundamental to medical practice.[1-3] Poor reasoning is an important cause of medical error; for example, diagnostic errors are thought to contribute to approximately 10% of patient deaths and hospital adverse events.[4] Most research in clinical reasoning has focused on decisions related to diagnosis, ie, diagnostic reasoning.

By contrast, management reasoning—which we define as the process of making decisions about patient management, including choices about treatment, follow-up visits, further testing, and allocation of limited resources—remains less well understood.[2,3] Paradoxically, management actually may be more important: diagnosis is only a means to an end (namely, proper management),[5] clinicians must frequently manage patients before making a definitive diagnosis, and diagnosis often hinges on management decisions (eg, choices regarding additional diagnostic testing). The distinction of diagnosis and management is not new; for example, informaticians commonly distinguish diagnostic support tools from drug prescribing tools. However, research in the cognitive processes of management reasoning is not well developed.

A better understanding of management reasoning could help to expand the current conception of medical

Features of Diagnostic and Management Reasoning

Diagnostic reasoning is primarily a classification task that assigns meaningful labels to a constellation of symptoms, examination findings, and test results.[1] These labels, which often connote a specific underlying pathology, shape the clinician's and patient's understanding of the underlying illness and facilitate communication among team members and with the patient.[5] By contrast, management reasoning is primarily a task of prioritization, shared decision making, and monitoring, and is typically more complex. Management reasoning differs from diagnostic reasoning in at least 5 ways.

First, diagnoses can be established as correct or incorrect. Different labels might appropriately be assigned to a given constellation of clinical findings (eg, "upper respiratory infection" or "acute sinusitis"), and for some diagnoses (eg, chronic Lyme disease) controversy exists regarding exactly what the label signifies. Nonetheless, the correctness of each label as an accurate reflection of the underlying illness can (at least in theory) be determined. Conversely, in management, the answer is often, "It depends." Patient preferences, societal values, logistical constraints, and resource availability appropriately influence management decisions. There are usually multiple paths to a successful outcome, and often multiple acceptable outcomes. Management reasoning involves contrasting, prioritizing, and selecting among the myriad reasonable (defensible) options.

Second, management plans are influenced by, and management reasoning must integrate, the preferences, values, resources, and constraints of the patient, clinician, other health care professionals, the institution, and payers.[6] Diagnostic reasoning, by contrast, is generally not influenced by values and preferences. A man with chest pain either is or is not experiencing a myocardial infarction regardless of his wishes or preferences, ability to pay, or proximity to a cardiac catheterization suite. The management plan (eg, whether to perform cardiac catheterization) depends on all these factors.

Third, diagnosis can often be done in isolation from the patient. A clinician can review recorded clinical information (history, examination findings, test results) and render a diagnosis; indeed, this is commonly done in solving clinical vignettes. Conversely, management decisions usually require communication and shared decision making with the patient and often with others including nurses, social workers, hospital administrators, insurance agencies, and public policy makers, each of whom could have an interest in how a condition is managed. Although multiple-choice questions frequently recognize only one "next best step in management," in reality, selecting the best option nearly always requires some element of negotiation.

Fourth, management plans are inherently fluid and require ongoing monitoring and frequent adjustments. Experienced clinicians can often anticipate such adjustments ("We will start with lifestyle measures; if

typically be definitively assigned at a given point in time. The diagnosis might subsequently change over time ("stage II prostate cancer" becomes "prostate cancer in remission" or "stage IV prostate cancer"), but this does not invalidate the correctness and completeness of the initial diagnosis.

Fifth, diagnostic reasoning involves a finite range of solutions and interacting factors. Additionally, some uncertainties in diagnosis (eg, incomplete information) can be accommodated by applying less-specific labels (eg, accepting a diagnosis of "community-acquired pneumonia" without trying to specify the causative pathogen). By contrast, management reasoning involves a dynamic interplay among people, systems, settings, and competing priorities, and is thus inherently complex and contextually "situated."[7] Moreover, management often entails more uncertainties (eg, unpredictable response to treatment) and these uncertainties often require a more detailed plan (ie, broader-based treatments, more contingencies, more frequent monitoring).

Research Priorities

Several areas of management reasoning require further research. First, management reasoning is taught, assessed, and then practiced under the presumption that it involves cognitive processes similar to diagnostic reasoning. This presumption should be tested. For example, experts commonly incorporate highly efficient and accurate pattern recognition processes in rendering a diagnosis.[2] Yet the explicit consideration of treatment costs and benefits, the use of rubrics to guide management decisions, and the integration of each patient's unique circumstances all suggest a deliberate, analytical process. As such, the balance among these cognitive processes may differ for management reasoning.

Second, better understanding is needed regarding how to define and recognize management errors. In contrast with diagnoses, for which a definitive correct/incorrect judgment can (at least in theory) usually be made, multiple reasonable management options nearly always exist.[7] It is also difficult to judge the correctness of patient preferences and how these are integrated; clinicians with effective yet different communication approaches might elicit different values or prioritize values differently, resulting in different management decisions. A good management plan might result in a poor clinical outcome (progression of cancer despite optimal therapy), a suboptimal plan might result in a good outcome (antibacterial treatment for viral respiratory infection), and a plan based on an incorrect or uncertain diagnosis might be correct for the diagnosis but suboptimal overall (aspirin for myocardial ischemia subsequently characterized as caused by coronary artery dissection). These issues might be simplified in nonclinical research settings by controlling key aspects of the clinical problem (eg, the information in a vignette or the specificity of a diagnosis), but in both controlled and clinical research settings the definition of management error requires careful consideration.

Third, research is needed to determine how to teach management reasoning, both in training and

monitoring of treatment response, yet how to optimally facilitate learning these and other management skills remains unknown. Clinicians must also learn to acknowledge complexity and act under conditions of uncertainty. Assessment of management reasoning must not only address these skills, but also use methods that accommodate multiple plausible solutions.

Fourth, strategies that support management reasoning in clinical practice need to be identified and implemented. In particular, tools and processes will be needed to help manage the interacting and situation-specific factors (ie, characteristics and preferences of the patient, clinical team, and health care system) that make management complex and potentially overwhelming. These strategies might include involving all members of the health care team in eliciting preferences, educating patients, articulating a plan, and monitoring therapy; implementing novel technologies to support management tasks (eg, computer-generated or crowdsourced treatment recommendations; tools that support clinicians in identifying and prioritizing multiple alternative plans); facilitating access to information that supports the integration of patient values and preferences (eg, decision aids explaining the patient-specific prognosis and benefits, risks, and out-of-pocket costs of potential test and treatment options); and monitoring the cognitive task load to permit intervention before it exceeds the clinician's cognitive capacity. Research and practice innovations have already identified both problems and potential solutions in this area[4]; viewing these issues through the distinct lenses of diagnostic and management reasoning may facilitate additional insights.

Article Information

Back to top

Corresponding Author: David A. Cook, MD, MHPE, Division of General Internal Medicine, Mayo Clinic College of Medicine, 200 First St SW, Mayo 17-W, Rochester, MN 55905 (cook.david33@mayo.edu).

Published Online: May 10, 2018. doi:10.1001/jama.2018.4385

Conflict of Interest Disclosures: All authors have completed and submitted the ICMJE Form for Disclosure of Potential Conflicts of Interest and none were reported.

Disclaimer: The views expressed herein are those of the authors and not necessarily those of the Department of Defense or other federal agencies.

Additional Contributions: We thank Larry Gruppen, PhD (Department of Learning Health Sciences, University of Michigan, Ann Arbor), for his substantial contributions in elaborating the ideas expressed in this article; he did not receive compensation for these contributions.

References

Academic Publishers; 2002:205-230.
Crossref

2. Norman G. Research in clinical reasoning: past history and current trends. *Med Educ*. 2005;39(4):418-427.
PubMed | Google Scholar | Crossref

3. McBee E, Ratcliffe T, Goldszmidt M, et al. Clinical reasoning tasks and resident physicians: what do they reason about? *Acad Med*. 2016;91(7):1022-1028.
PubMed | Google Scholar | Crossref

4. Balogh EP, Miller BT, Ball JR, eds; National Academies of Sciences, Engineering, and Medicine. *Improving Diagnosis in Health Care*. Washington, DC: National Academies Press; 2015.
Crossref

5. Ilgen JS, Eva KW, Regehr G. What's in a label? is diagnosis the start or the end of clinical reasoning? *J Gen Intern Med*. 2016;31(4):435-437.
PubMed | Google Scholar | Crossref

6. Mercuri M, Sherbino J, Sedran RJ, Frank JR, Gafni A, Norman G. When guidelines don't guide: the effect of patient context on management decisions based on clinical practice guidelines. *Acad Med*. 2015;90(2):191-196.
PubMed | Google Scholar | Crossref

7. Durning SJ, Artino AR Jr, Schuwirth L, van der Vleuten C. Clarifying assumptions to enhance our understanding and assessment of clinical reasoning. *Acad Med*. 2013;88(4):442-448.
PubMed | Google Scholar | Crossref

See More About

Oncology | Prostate Cancer | Urologic Cancer | Urology | Clinical Decision Support

Others Also Liked

Clinicians' cognitive biases: a potential barrier to implementation of evidence-based clinical practice

Claudia Caroline Dobler et al., Evid Based Med, 2019

Primary care clinicians' views of paediatric respiratory infection surveillance information to inform clinical decision-making: a qualitative study

Emma C Anderson et al., Paediatrics Open, 2019

Trending

Effect of Teaching Bayesian Methods on Medical Students' Ability to Estimate Probability of a Diagnosis

JAMA Network Open | *Research* | *December 20, 2019*

Adolescent Cognitive Aptitudes and Later-in-Life Alzheimer Disease and Related Disorders

JAMA Network Open | *Research* | *September 7, 2018*

Genetic Loci Influencing Schizophrenia Risk and Cognitive Traits

JAMA Psychiatry | *Research* | *October 1, 2017*

Examining the Patterns of Uncertainty Across Clinical Reasoning Tasks: Effects of Contextual Factors on Clinical Reasoning Performance

Short title: Uncertainty in Clinical Reasoning Tasks

Divya Ramani, MS, Michael Soh, PhD, Jerusalem Merkebu, PhD, Steven J. Durning, MD, PhD, Alexis Battista, PhD, Elexis Mcbee, DO, Temple Ratcliff, MD, Abigail Konopasky, PhD

Corresponding Author:

Divya Ramani

Research Associate

Uniformed Services University

4301 Jones Bridge Road

Building 53-R30

Bethesda, MD

Word Count: 3500
Number of Tables: 3
Number of Figures: 1
Supplemental Material: No

Abstract:

Introduction

Uncertainty is common in clinical reasoning given the dynamic processes required to come to a diagnosis. Though some uncertainty is expected during clinical encounters, it can have detrimental effects on clinical reasoning. Likewise, evidence has established the potentially detrimental effects of the presence of distracting contextual factors (i.e., factors other than case content needed to establish a diagnosis) in a clinical encounter on clinical reasoning. The purpose of this study was to examine how linguistic markers of uncertainty overlap with different clinical reasoning tasks and how distracting contextual factors might affect physicians' clinical reasoning process.

Materials and Methods

In this descriptive exploratory study, physicians participated in a live or video recorded simulated clinical encounter depicting a patient with unstable angina with and without contextual factors. Transcribed think-aloud reflections were coded using Goldszmidt's clinical reasoning task typology (26 tasks encompassing the domains of framing, diagnosis, management, and reflection) and then those coded categories were examined using linguistic markers of uncertainty (e.g., probably, possibly, etc.).

Results

Thirty physicians with varying levels of experience participated. Consistent with expectations, descriptive analysis revealed that physicians expressed more uncertainty in cases with distracting contextual factors compared to those without. Across the four domains of reasoning tasks, physicians expressed the most uncertainty in diagnosis and least in reflection.

Conclusion

These results highlight how linguistic markers of uncertainty can shed light on the role contextual factors might play in uncertainty which can lead to error and how it is essential to find ways of managing it.

Keywords: *Context specificity, contextual factors, uncertainty, clinical reasoning, clinical reasoning tasks, linguistics*

Examining the Patterns of Uncertainty across Clinical Reasoning Tasks: Effects of Contextual Factors on Clinical Reasoning Performance

Introduction

Clinical reasoning is best conceptualized as the steps up to and including arriving at a diagnosis and devising a treatment plan based on the integration of information derived from the clinical encounter (e.g., patient interview, physical exam findings, etc).[1,2] Clinical reasoning is an emergent process that is jointly constructed by the patient, physician and other participants in the encounter, and thus is influenced by *contextual factors* (factors arising from patient, physician and environment like language barriers or misleading diagnostic suggestion).[3,4] When contextual factors are present, physicians may arrive at two different diagnoses for two patients with identical symptoms and findings who have the same diagnosis, a phenomenon called *context specificity*.[4–6] In this way, the presence of contextual factors can lead to diagnostic error.

Situated cognition theory argues that learning and performance are shaped by and inseparable from the contexts of human behavior, cultural and social practices (e.g., clinical practices), and language.[7] Situated cognition can help illuminate the phenomenon of context specificity by emphasizing how participants (e.g., physician and patient), environment (e.g. availability or absence of medical resources), and linguistic production (e.g., conversation) are part of the larger processes that shape clinical reasoning during a patient encounter.[4,7] Rather than viewing clinical reasoning solely a linear series of internal decisions, situated cognition conceptualizes reasoning as emerging dynamically from the specifics of the situation. Thus, situated cognition is an imperative lens for understanding *how* contextual factors such as diagnostic suggestion, patient language barriers, physician burnout, limited encounter time, and

lack of medical resources, among others, can affect diagnostic reliability and hamper patient safety.[4,6,8]

According to Goldszmidt and colleagues, physicians engage in distinct clinical reasoning *tasks* as they work towards a diagnosis and treatment.[9–11] Based on interviews with experts, Goldszmidt and colleagues developed a unified list of 24 clinical reasoning tasks--expanded in later applied work to 26--which physicians engage in during a clinical encounter.[9–11] These tasks are divided into four domains: a) framing (e.g., identifying active issues), b) diagnosis (e.g., identifying risk factors), c) management (e.g., establishing goals of care), and d) reflection (e.g., identifying knowledge gaps).[9,11] While recent work has begun to describe patterns in the use of these tasks, no study has yet used the framework of these tasks to examine physician uncertainty and how that may shift depending upon the unique context.

Reasoning through and having to make decisions based on numerous interwoven factors can lead physicians to feelings of uncertainty.[12,13] Uncertainty, "an awareness of incomplete understanding of a situation or event" (p. 866), manifests sometimes in clinical reasoning as difficulty determining diagnosis and treatment plans.[14] Thus, it has become a topic of increasing interest in medicine where efforts are focused on understanding uncertainty's role in clinical encounters as well as addressing ways of overcoming it.[15] Among other sources, uncertainty in clinical reasoning arises from case complexity or ambiguity, a lack of information or experience with a specific case, and the complex and emergent relationship between patient and physician.[13–15]

Beyond the clinical reasoning tasks themselves, uncertainty can be introduced by various contextual factors associated with the physician, patient, and environment.[5] These factors can increase uncertainty by interfering with the collection of appropriate evidence (e.g., not asking

about certain symptoms due challenges with processing information provided) or with the use of that evidence to make a diagnosis (e.g., anchoring on an "obvious" diagnosis due to uncertainty about what other diseases could cause the presentation).[16,17] Taken together, examining uncertainty markers within the Goldszmidt framework[11] of clinical reasoning tasks may help us to better understand where in the clinical reasoning process uncertainty emerges and how, if at all, the presence of contextual factors changes this pattern. This will allow us to appropriately support physicians in their uncertainty across a variety of contexts.

Thus, this study examines patterns of uncertainty and clinical reasoning tasks in cases with and without potentially distracting contextual factors. More specifically, we asked,

1. Does frequency of uncertainty markers differ in cases with and without contextual factors?
2. Do patterns of uncertainty markers across framing, diagnosis, management, and reflection differ in cases with and without contextual factors? If so, how?
3. How, if at all, does the use of uncertainty in diagnostic tasks differ in cases with and without contextual factors?

Materials and Methods

This research is situated within a larger body of research using situated cognition (among other theoretical constructs) to examine physicians' clinical reasoning performance.

Sample Selection

Based on findings from the larger study, we identified a sample of participant performances in simulated clinical encounters of practicing physicians (either intermediate, with 10 years' or less experience, and *experienced*, with over 10 years' experience) in internal medicine, family medicine, and surgery. Because participants experienced more difficulty with

one of the encounters (an unstable angina case), we purposefully sampled each participant's unstable angina case for this study, comparing those participants who had an unstable angina case with distracting contextual factors to those whose case did not have distracting contextual factors.[16]

Procedure Participants were quasi-randomly assigned to either a video or live simulation condition. Participants in both conditions were asked to either view a pre-recorded video depicting a clinical encounter or engage in a live scenario-based simulation with a standardized patient and then provide a diagnosis and treatment plan. The case content along with the distracting contextual factors was controlled for both conditions. The chosen contextual factors were based on common occurrences in clinical practice (for the angina case they were: the patient offering a misleading diagnostic suggestion and the patient reporting history in circuitous manner. Immediately following participation in the live or video encounter, participants were asked to "think aloud" while they either rewatched the video or watched their own video performance, talking about how they arrived at the diagnosis.[18,19] This think-aloud method has been a valuable tool in prior studies to explore clinical reasoning.[19,20] The study protocol was approved by the Institutional Review Board at the Uniformed Services University in Bethesda, Maryland.

Data Analysis

Analysis of transcribed think-alouds for this study was conducted in four stages.

Stage I: Task-based coding. To support our goal of examining uncertainty within clinical reasoning tasks, we first coded each transcript drawing on Goldzsmidt and colleagues'

four domains: a) framing (made up of three tasks, including: identifying active issues), b) diagnosis (made up of eight tasks, including: prioritizing differential diagnosis and selecting diagnostic investigations), c) management (made up of 13 tasks, including: establishing goals of care and assessing illness severity), and d) reflection (made up of 2 tasks: considering cognitive bias and identify knowledge gaps).[11] Two physicians (EM & TR) used this coding schema to code participant think aloud for these clinical reasoning tasks. The tasks were coded separately and then reviewed together to arrive at consensus.

Stage II: Uncertainty coding. To examine patterns of uncertainty, the think alouds that were coded for clinical reasoning tasks were also coded for linguistic markers of uncertainty (e.g. probably, possibly, etc.), that have been identified in prior work in medicine to identify patterns of uncertainty. [16,21] Three team members (DR, MS, JM) used this coding schema and then reviewed as a group to arrive at consensus. The number of uncertainty markers for each of a participant's two cases were then recorded.

Stage III: Inferential and exploratory data analysis: Frequencies and percentages of uncertainty markers across all four clinical reasoning domains, with and without contextual factors, were calculated. Additionally, an independent samples t-test was conducted to compare the rate of uncertainty in the presence and absence of contextual factors.

Stage IV: Qualitative follow-on analysis: Based on results of inferential and exploratory quantitative analysis, we conducted a comparative thematic analysis of the pattern of uncertainty markers. For each of the 26 tasks, we compared instances of uncertainty markers in cases with and without contextual factors, seeking to categorize what aspect of the clinical situation (e.g., patient symptoms, participant clinical knowledge, etc) participants were uncertain

about and whether these patterns of uncertainty *differed* in cases with and without contextual factors.

Results

Participants were 30 physicians (11 women, 19 men) from internal medicine (n =22), family medicine (n =3), and surgery (n = 5), with varying levels of experience (21 intermediate physicians, 10 years or less experience; 9 experienced physicians, over 10 years' experience). The unstable angina case of 20 participants had distracting contextual factors and the unstable angina case of 10 participants had no distracting contextual factors. Only one of the 30 transcripts did not have uncertainty markers and we excluded this outlier (which had no contextual factors) from further analysis. Overall, transcripts were coded for a total of 335 uncertainty markers (see Table 1 for examples) and 1117 clinical reasoning tasks (see Table 2 for examples). Intermediate physicians had more uncertainty markers (23% of clinical reasoning tasks had uncertainty markers) than experienced physicians (only 12% of tasks had uncertainty markers).

Quantitative Results

Overall, physicians expressed a higher rate of uncertainty when in the presence of a contextual factor (31% of clinical reasoning tasks coded in a contextual factor case had uncertainty markers) than not (27% of clinical reasoning tasks in a non-contextual factor case had uncertainty markers). Independent sample t-test analysis revealed this difference to be non-significant (t (457.65) = 1.22, p = 0.225) with low practical significance (d = 0.09). Subsequently, analysis of the four types of clinical reasoning tasks (framing, diagnosis, management, and reflection) revealed that physicians express uncertainty most during diagnosis (70% of uncertainty markers fall in diagnostic tasks), followed by framing (17% fall in framing

tasks), then management (11% fall in management tasks) and least in reflection (2% fall in reflection tasks; see Table 3). As Table 3 indicates, this distribution of uncertainty markers across types of clinical reasoning tasks is relatively similar with and without distracting contextual factors (e.g., 16% of uncertainty markers occur in framing tasks in the presence of contextual factors and 21% occur in framing tasks in the absence of contextual factors).

Qualitative Results

In order to better understand this potential trend in results between the contexts, we qualitatively compared the *use* of uncertainty markers in clinical reasoning tasks between both contexts, focusing on the diagnostic tasks (tasks 4 through 11). First, for the critical task of determining the most likely diagnosis and underlying causes (task 7), only two participants denote uncertainty in the absence of contextual factors, versus five in their presence. Of those five participants, two use uncertainty markers to pose what are incorrect diagnoses ("It's **probably**, **likely**, pericarditis" and "reflux is probably the most likely thing"). Similarly, for the diagnostic task of considering and prioritizing differential diagnoses (task 4), participants in the contextual factor condition offer a wider range of possibilities of what the diagnosis "could" or "might" be (e.g., nicotine withdrawal, congestive heart failure, panic attack etc). Thus, the presence of contextual factors in this sample seemed to elicit a broader range of leading and differential diagnoses, which in some scenarios may be beneficial.

Second, for the tasks that involved generating underlying causes (identifying precipitants or triggers to the current problem [task 5] and identifying modifiable and non-modifiable risk factors [task 8], participants express uncertainty around a wider variety of causes in the presence of contextual factors (e.g., caffeine or an energy supplement, a potential drinking habit).

Meanwhile, in the non-contextual factor condition, participant uncertainty focuses on a narrower range of potentially relevant features like age, smoking history, and family history. Moreover, when generating reasoning processes like these when contextual factors are present, participants are more uncertain about their *own actions* versus the processes themselves (e.g., "I **should** have asked about, **like**, life stressors, work stressors").

Third, regarding selection of diagnostic investigations (task 6), we found 15 uses of uncertainty markers in the contextual factors condition and only *one* in the non-contextual factors condition. In the presence of contextual factors, participants speculated uncertainly about a wide variety of potential diagnostic tests: "maybe" a TB workup; he "may" need an endoscopy; or "to "kind of get a baseline." While participants in both conditions suggested diagnostic investigations, there was more uncertainty around a wider *range* of them when contextual factors were present.

The other three diagnostic tasks—identifying complications associated with diagnosis or treatment (task 9), assessing rate of progression, response to treatment, and prognosis (task 10), and exploring physical and psychosocial consequences of treatment (task 11)—occurred sporadically in the data, but are *only* used with uncertainty markers in the presence of contextual factors. Of the 11 instances of these three tasks, four of them related to uncertainty about *patient* behavior or reporting, not the diagnosis, disease progression, or treatment consequences themselves (e.g., "if he does have uncontrolled blood pressure, **maybe** he isn't taking his medications").

Discussion

In this paper, we examined patterns in uncertainty markers while physicians engaged in framing, diagnostic, management, and reflection clinical reasoning tasks throughout a case of

unstable angina with and without distracting contextual factors. The findings revealed that physician uncertainty trended higher, though not statistically significantly, in the presence of distracting contextual factors than in their absence, which is in line with our previous research.[4,16] Thus, we also conducted a qualitative analysis to better understand the differences between expressions of uncertainty with and without contextual factors.

While all clinical encounters inherently have a certain degree of uncertainty, these findings suggest that contextual factors can introduce an additional level of ambiguity or complexity that may impede the reasoning process, creating even more uncertainty.[22] Another possible reason for increased uncertainty is that, as argued in prior work, contextual factors may increase cognitive load, which constrains the use of working memory.[4,5,8] When the ability to make full use of working memory is hampered, it may further introduce uncertainty as the physician has less capacity to process the wealth of other information present in a clinical encounter. Furthermore, our findings indicated that intermediate physicians (with less than 10 years of work experience) exhibited more uncertainty than experienced physicians in the presence of distracting contextual factors. This further supports past research contending that physicians with less experience have lower tolerance towards uncertainty and, hence, are less able to manage it.[22,23]

Our examination of the distribution of these uncertainty markers across the four clinical reasoning task types, however, did *not* reveal differences in cases with and without contextual factors. Looking more closely at each of the eight diagnostic tasks qualitatively, however, uncertainty *generates* emerge as a difference in the range of options was observed. The presence of contextual factors was associated with uncertainty about a wider variety of potential diagnoses, underlying causes, and diagnostic tests. One effect of context specificity, then,

appears to be the creation of conditions for positing a broader range of clinical *possibilities*, whether the possibility be a diagnosis like costochondritis, a trigger like an energy supplement, or investigations like a TB workup. Physicians in the contextual factors condition were perhaps holding in their minds a wider variety of potential diagnostic and treatment paths, which could potentially contribute to increased cognitive load.[5] This increase in diagnostic and treatment paths may be beneficial for some patients, but not others.[24] As such, considering a wider range of possibilities could improve or hamper patient care and thus, requires context sensitivity.

Participants in contextual factors cases also appeared to exhibit more uncertainty around patient behavior and information derived from their reporting than participants in non-contextual factor cases. Thus, not only are they considering more diagnostic and treatment options, but they may also be debating whether the patient is a trustworthy reporter; yet another source of cognitive load that may prove counterproductive.

Our study has several limitations. First, our sample size is small and groups were unevenly distributed. However, as an exploratory study, it does raise concerns about uncertainty and clinical reasoning and the need for further studies of uncertainty in medicine and how to mitigate it. Second, participants were distributed across the video and live simulation conditions. Future studies might focus on a single simulation modality as this may have an effect on the way physicians reason clinically.

This work has several important implications. First, the available evidence suggests that uncertainty is influenced by context and task which is consistent with situated cognition theory and the notion that clinical reasoning is an emergent phenomenon. Our work suggests that developing metacognitive awareness of patient, physician, and environmental factors in the presence of uncertainty could improve patient care as it seems uncertainty is related to high

cognitive load, specific instruction and awareness of these factors. Our work also suggests that there are myriad of reasons why uncertainty occurs and that a “one size fits all” approach is unlikely to be beneficial. Finally, there are times that uncertainty may improve clinical care by prompting the physician to generate, for example, an expanded differential diagnoses. Whether this is a benefit or hindrance and how it unfolds in clinical reasoning should be explored in future studies.

Given the ubiquity of uncertainty in clinical reasoning and its potential influence on reasoning, an important research endeavor would be to expand upon the work of Goldszmidt and colleagues,[9–11] exploring both the tolerance for and management of uncertainty as a clinical reasoning task. We hope this work initiates important research and practice discussions on how to better address uncertainty and its role in the clinical reasoning process.

Acknowledgements: The authors would like to thank Megan Ohmer, for her help in carefully transcribing the think aloud.

Author Contributions: All authors collaborated together on the research design and analysis of the data. All authors offered substantive revisions and approve of this final version of the paper.

Research Funding: This study was supported by a grant from the JPC - 1, CDMRP - Congressionally Directed Medical Research Program (#NH83382416)

Employment or Leadership: None Declared

Competing Interest: The funding organization played no role in the study design; in the collection, analysis, and interpretation of data; in the writing of the report; or in the decision to submit the report for publication.

References

1. Young M, Thomas A, Lubarsky S, et al. Drawing Boundaries. *Acad Med*. 2018:1. doi:10.1097/ACM.0000000000002142
2. ten Cate O, Durning SJ. Understanding Clinical Reasoning from Multiple Perspectives: A Conceptual and Theoretical Overview. 2018:35-46. doi:10.1007/978-3-319-64828-6_3
3. Durning S, Artino Jr AR, Pangaro L, van der Vleuten CPM, Schuwirth L. Context and clinical reasoning: understanding the perspective of the expert's voice. *Med Educ*. 2011;45(9):927-938.
4. Durning SJ, Artino AR, Boulet JR, Dorrance K, van der Vleuten C, Schuwirth L. The impact of selected contextual factors on experts' clinical reasoning performance (does context impact clinical reasoning performance in experts?). *Adv Heal Sci Educ*. 2012;17(1):65-79.
5. Ratcliffe TA, McBee E, Schuwirth L, et al. Exploring Implications of Context Specificity and Cognitive Load in Residents. *MedEdPublish*. 2017;6.
6. Eva KW. What every teacher needs to know about clinical reasoning. *Med Educ*. 2005;39(1):98-106.
7. Brown JS, Collins A, Duguid P. Situated cognition and the culture of learning. *Educ Res*. 1989;18(1):32-42.
8. McBee E, Ratcliffe T, Picho K, et al. Consequences of contextual factors on clinical reasoning in resident physicians. *Adv Heal Sci Educ*. 2015;20(5):1225-1236.
9. Goldszmidt M, Minda JP, Bordage G. Developing a unified list of physicians' reasoning tasks during clinical encounters. *Acad Med*. 2013;88(3):390-394.
10. McBee E, Ratcliffe T, Goldszmidt M, et al. Clinical reasoning tasks and resident

physicians: what do they reason about? *Acad Med*. 2016;91(7):1022-1028.

11. Juma S, Goldszmidt M. What physicians reason about during admission case review. *Adv Heal Sci Educ*. 2017;22(3):691-711.
12. Hillen MA, Gutheil CM, Strout TD, Smets EMA, Han PKJ. Tolerance of uncertainty: conceptual analysis, integrative model, and implications for healthcare. *Soc Sci Med*. 2017;180:62-75.
13. Wakeham J. Uncertainty: history of the concept. 2015.
14. Dhawale T, Steuten LM, Deeg HJ. Uncertainty of physicians and patients in medical decision making. 2017.
15. Han PKJ, Klein WMP, Arora NK. Varieties of uncertainty in health care: a conceptual taxonomy. *Med Decis Mak*. 2011;31(6):828-838.
16. Konopasky AW, Ramani D, Ohmer M, et al. It Totally Possibly Could Be: How A Group Of Military Physicians Reflect On Their Clinical Reasoning In The Presence Of Contextual Factors. *Mil Med*.
17. Surry LT, Torre D, Trowbridge RL, Durning SJ. A mixed-methods exploration of cognitive dispositions to respond and clinical reasoning errors with multiple choice questions. *BMC Med Educ*. 2018;18(1):277.
18. Elstein AS, Shulman LS, Sprafka SA. Medical problem solving an analysis of clinical reasoning. 1978.
19. Ericsson KA, Simon HA. How to study thinking in everyday life: Contrasting think-aloud protocols with descriptions and explanations of thinking. *Mind, Cult Act*. 1998;5(3):178-186.
20. Durning SJ, Artino AR, Beckman TJ, et al. Does the think-aloud protocol reflect thinking?

Exploring functional neuroimaging differences with thinking (answering multiple choice questions) versus thinking aloud. *Med Teach*. 2013;35(9):720-726. doi:10.3109/0142159X.2013.801938

21. Bhise V, Rajan SS, Sittig DF, Morgan RO, Chaudhary P, Singh H. Defining and measuring diagnostic uncertainty in medicine: a systematic review. *J Gen Intern Med*. 2018:1-13.
22. Iannello P, Mottini A, Tirelli S, Riva S, Antonietti A. Ambiguity and uncertainty tolerance, need for cognition, and their association with stress. A study among Italian practicing physicians. *Med Educ Online*. 2017;22(1):1270009.
23. Lawton R, Robinson O, Harrison R, Mason S, Conner M, Wilson B. Are more experienced clinicians better able to tolerate uncertainty and manage risks? A vignette study of doctors in three NHS emergency departments in England. *BMJ Qual Saf*. 2019;28(5):382-388.
24. DeKay ML, Asch DA. Is the defensive use of diagnostic tests good for patients, or bad? *Med Decis Mak*. 1998;18(1):19-28.

Table 1: Examples from think alouds reflecting patterns of uncertainty

	Examples
1	"You know, my leading diagnosis would be ***probably*** angina"
2	"He says ***maybe*** it is faster during these episodes of chest pain"
3	"so ***kind of*** leading to ***maybe*** a cardiac, um, cause of this pain"
4	"have to ***sort of*** exclude heart disease or… until you think about anything else"
5	"***perhaps*** if he was having panic attack"
6	"then I would be calling cardiology to find out when they ***could*** be getting him in for a stress test"

Table 2: Examples from "Think alouds" reflecting the four domains of task based clinical reasoning

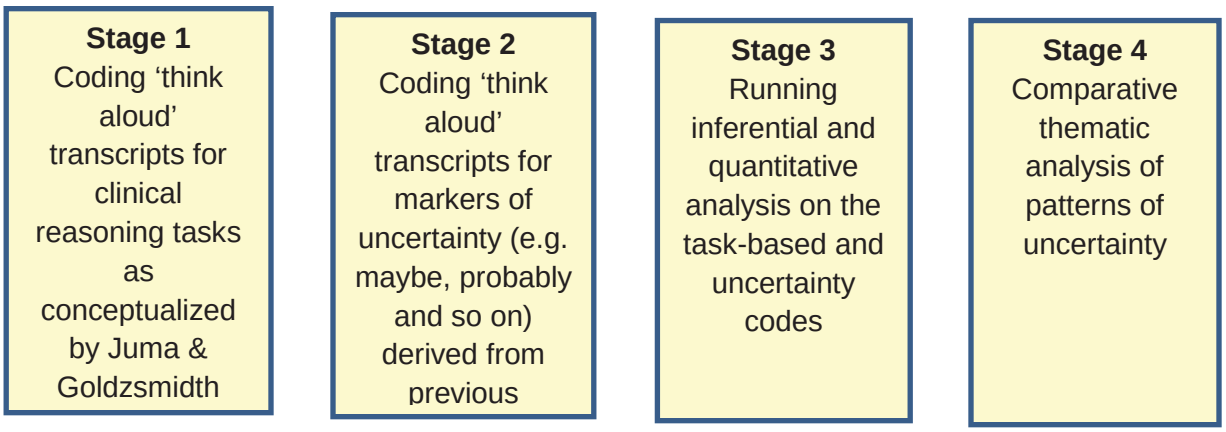
Clinical Reasoning Task	Examples
Framing	"There's no swelling or pain in his legs"
Diagnosis	"I wanted to rule out, maybe, pulmonary embolism, anything like that, that could be contributing"
Management	"I just want to explain to him what's going on and what I'm thinking because he's clearly, probably already thinking about it. So, just acknowledging it and making sure that he gets a chance to answer any questions about the possibility".
Reflection	"I noticed that I asked a few leading questions, getting into it and I feel that I should have just asked 'pattern' first and then, if he needed prompting then go to, like, 'For example if…'"

Table 3: Overall emergence of uncertainty in task based reasoning (per think aloud)

	Framing	Diagnosis	Management	Reflection

Percent uncertainty with contextual factors	16%	68%	14%	3%
Percent uncertainty, no contextual factors	21%	77%	3%	0
Percent uncertainty across all transcripts	17%	70%	11%	2%

Figure 1: Stages of Analysis



ACADEMIC MEDICINE
Journal of the Association of American Medical Colleges

Author: Samuel Anita PhD; Konopasky Abigail PhD; Schuwirth Lambert W.T. MD, PhD; King Svetlana M. PhD; Durning Steven J. MD, PhD

Title: Five Principles for Using Educational Theory: Strategies for Advancing Health Professions Education Research

DOI: 10.1097/ACM.0000000000003066

Academic Medicine

DOI: 10.1097/ACM.0000000000003066

Five Principles for Using Educational Theory: Strategies for Advancing Health Professions Education Research

Anita Samuel, PhD, Abigail Konopasky, PhD, Lambert W.T. Schuwirth, MD, PhD, Svetlana M. King, PhD, and Steven J. Durning, MD, PhD

A. Samuel is assistant professor, Graduate Programs in Health Professions Education, Uniformed Services University, Bethesda, Maryland.

A. Konopasky is assistant professor, Graduate Programs in Health Professions Education, Uniformed Services University, Bethesda, Maryland.

L.W.T. Schuwirth is professor and director, Prideaux Centre for Research in Health Professions Education, Flinders University, Adelaide, South Australia, Australia.

S.M. King is lecturer, Prideaux Centre for Research in Health Professions Education, Flinders University, Adelaide, South Australia, Australia.

S.J. Durning is professor and director, Graduate Programs in Health Professions Education, Uniformed Services University, Bethesda, Maryland.

Correspondence should be addressed to Anita Samuel, Uniformed Services University, Department of Medicine, Graduate Programs in Health Professions Education, 4301 Jones Bridge Road, Bethesda, MD 20814; telephone: 414-678-8847; email: anita.samuel.ctr@usuhs.edu.

Funding/Support: None.

Other disclosures: None reported.

Ethical approval: Not applicable.

Disclaimers: The opinions and assertions expressed herein are those of the authors and do not necessarily reflect the official policy or position of the Uniformed Services University or the Department of Defense.

Abstract

Health professions education (HPE) research often involves examining complex phenomena. Theory provides a means for better understanding the mechanics of these phenomena and guiding health professions researchers and educators as they navigate the practical implications for teaching, learning, and research. Engaging with educational theory is, therefore, critical to facilitating this understanding. However, this engagement presents a key challenge for HPE researchers and educators without a background in social science. This article outlines 5 key principles of engaging with theory and offers integration strategies to assist HPE researchers and educators who wish to apply theory to their HPE scholarship and practice. The article concludes with a practical example of how these principles were applied to an HPE research project, demonstrating the value of theory in enhancing research quality. Existing theories can facilitate opportunities for individual researchers to better understand complex phenomena while simultaneously moving forward the field of HPE.

The health professions education (HPE) literature is relatively new and is growing and diversifying.[1] Educators and researchers in HPE value the input of scholars from diverse backgrounds who inform the work of HPE. Indeed, this diversity, coupled with the ability to empirically study how theory "works" in practice, is a strength of HPE that sets it apart from other fields[2] of enquiry. An additional point of distinction is that HPE scholars face complex challenges that have an impact on the education of learners and the health care of nations, for example, rapid innovations in technology in health care and education with fundamental societal impacts. In HPE, we use theory to better understand the mechanics of the phenomena underlying these complex challenges and to guide us as we navigate the practical implications of these challenges. However, HPE researchers and educators, particularly those who do not have a background in education or social sciences, may lack a strong grasp of theory.[2]

Using theory in HPE is a balancing act, as Ostrom notes:

> Without theory, one can never understand the general underlying mechanisms that operate in many guises in different situations. If not harnessed to solving empirical puzzles, theoretical work can spin off under its own momentum, reflecting little of the empirical world.[3]

Overreliance upon theory runs the risk of jeopardizing the work's connection with educational practice. Conversely, if HPE research remains atheoretical (purely practical), there is a risk of pursuing inefficient and ineffective educational innovations. Understanding theory is also important in preventing a reinvention of the wheel. For example, with the advent of new technology, innovation is sometimes implemented without a deeper theoretical understanding of why previous educational innovations were ineffective. Consider the use of new technologies to assess the "skill" of clinical reasoning, for example. Moving from clever, paper-based solutions

to virtual reality solutions ignores the fact that the issues are not a matter of fidelity but, rather, are more fundamental to the nature of clinical reasoning (i.e., superior performance is not solely based on some sort of generic problem-solving ability).[4]

To incorporate theory effectively, we must understand its scope and limitations. In this article, we have outlined 5 key principles for using theory. We have also suggested strategies for those who wish to integrate theory into their work to improve the scholarship and practice of HPE.

Principle 1: All Theories Are Not Created Equal

Theories are a way to decontextualize or recontextualize complex educational problems to enhance our understanding and to guide potential courses of action. They provide lenses for viewing a situation or problem, and not all lenses are equal. Some theories provide a general structure or scaffold to view a problem. Situated cognition is an example of such a theory. It argues that learning is strongly influenced by the situation in which it occurs and directs the researcher to examine the components of that situation and how they interact.[5] In so doing, aspects of the situation and connections among those aspects may emerge that would otherwise have remained unexamined had this theory not been applied. The result is a better understanding of what makes learning effective (or ineffective). Another example of a theory that provides a more general scaffolding is ecological psychology. It views a situation[6] (e.g., an educational activity) as a series of affordances (i.e., what the properties of the educational method allow or disallow the teacher and/or learners to do) and effectivities (i.e., those affordances that teachers and learners perceive to be in the educational method). Such "grand" theories[7,8] can be considered as macro-level theories because they attempt to theorize all aspects of a learning situation. This type of theory can be applied to many situations to help identify and understand the features that contribute to the educational processes and outcomes of interest. By their nature,

however, they often lack utility in terms of predicting what will happen in a given situation or of determining the best next steps. Consequently, these theories often cannot be tested in a single study but rather require a program of research (i.e., multiple studies pursuing a coherent or connected line of enquiry).[9]

Other theories are micro level[8] and can only be applied to specific circumstances and/or components of the learning environment. They offer a more microscopic view, often providing more explanatory and, at times, predictive power than the grand theories. Generalizability theory[10] is one example. This theory does not aim to explain multiple phenomena in complex situations but is a highly practical approach to separate components of variance in an assessment results matrix. In other words, once a design for a study has been determined, generalizability theory allows researchers to gauge potential error sources. While macro theories, such as situated cognition, help researchers to scaffold and enhance understanding of broader phenomena, micro theories, such as generalizability theory, enable researchers to work with data to create specific, functioning, defensible claims and inferences.

While combining micro and macro theories may seem counterintuitive, this approach is not dissimilar to practices in medicine. For instance, we use macro screening tests that are typically quite sensitive and then perform follow-up confirmatory (micro) tests that are more specific to the condition being considered. When we select an appropriate theory (i.e., macro, micro, or a combination of both), we are better able to align our theoretical perspective(s) with the problem under investigation. Although we have chosen micro and macro theories as illustrative examples, these perspectives do not exist as a dichotomy but, rather, along a continuum. Theories that provide macro perspectives are particularly useful for obtaining large-scale views of what is

occurring, while micro theories are helpful to incorporate into subsequent experiments. We will provide an example of how we combine theories later in the paper.

Principle 2: Multiple Theories Can Be Used in a Given Research Study

There are times when a single theory may be insufficient to address the research question posed. In these situations, multiple theories can (and should) be used to explore the phenomenon under investigation from different perspectives. When researchers are trying to understand a complex phenomenon such as clinical reasoning, for instance, cognitivist theories[11] may help to frame how reasoning incorporates learning experiences into long-term memory. In parallel, clinical reasoning can also be examined by using social constructivist theories.[5,6] This approach can aid in understanding how the quality of reasoning—judged by, for example, the clarity, coherence, and plausibility of the claims and evidence—promotes interaction between learner and teacher or peers and thus affects the quality of the learning process. Just as researchers might bring together qualitative interview data and quantitative assessment data to more fully understand both the processes and outcomes of a given educational intervention, they can use multiple theories to understand different aspects and perspectives of an HPE context.

Again, this approach is similar to the practice of interdisciplinary health care. Understanding the complex phenomenon of chronic, benign low back pain requires more than a single theoretical approach. Gaining a full understanding of all the facets of this medical problem may require consideration of neurological, psychological, orthopedic, physiotherapeutic, occupational, social, and pharmaceutical perspectives. Similarly, in HPE, we encourage scholars to consider how multiple theories can assist them with their work. This form of triangulation, at the theory level, can deepen our understanding of the educational problem. This triangulation is particularly relevant because of the nature of HPE as a field that often deals with complex phenomena.

There is, however, one caveat to this principle: The different theories should belong to the same domain[8] or paradigm of education.[12] If there is misalignment with the epistemological assumptions between theories (e.g., a view of knowledge as concrete and fixed versus a view of knowledge as malleable and shifting), it may prove very difficult to integrate study findings and generate comprehensible claims without an appropriate mixed methods framework.[13]

Principle 3: You Can Deviate From a Theory's Propositions

As discussed earlier, HPE is an interdisciplinary field of enquiry. Because of this and because HPE is relatively young (when compared with fields like psychology or sociology), existing theories are not always exclusively designed for our community. This requires researchers to remain agile in their application and use of theories. Consequently, we encourage HPE scholars not only to use existing theoretical propositions in their work but also to be open to revising theory as needed for better application to HPE settings. Good doctoral dissertations in HPE often require revising theories to facilitate better alignment to the topic and/or proposing an alternative theoretical model that uses more than one theory to explain the phenomenon under investigation. One example is using the element of transfer (from cognitivist theories)[14] to understand the development of assessment expertise, such as in the domain of rater decision-making processes.[15–16]

Challenging the prevailing treatment approaches is not uncommon in clinical practice either. Our approach to CPR, for instance, has changed dramatically over the course of a decade, with harder and faster compressions, a different ordering of steps (circulation first), and a new ratio of compressions to breaths. As HPE researchers use theories from other fields, they may need to adapt and modify them to better align with the learners, instructors, and environments found in the health professions.

Principle 4: Terminology Can Be Reconciled Across Theories

A challenge we face as HPE scholars is understanding the range of theories that can be applied to our growing field. We can facilitate interdisciplinary collaboration and enhance theoretical agility in the HPE community by clarifying terminology and simultaneously reducing jargon. One way to do so is by examining the similarities and differences in terms from different theories and to use the results of this comparison to revise the terminology for the problem under investigation. For example, there has been some confusion regarding the terms *self-regulation*, *self-direction*, and *self-determination*.[17] One approach would be to avoid the conundrum and to choose a single theory, say self-regulated learning, and to adhere to its terminology. In so doing, however, the researcher loses potential insights on adult learning that are specific to self-directed learning theory and on intrinsic motivation that are specific to self-determination theory. Instead, explicitly and carefully comparing and revising these terms can help make the associated theoretical contexts more transparent and better support the specific work (e.g., nursing students studying for exams, interprofessional teams working on communication strategies, or practitioners improving their ongoing practice). By enhancing our understanding of terminology—and identifying the similarities and differences across theories—we can develop an agile toolkit to help HPE scholars move the field forward.

Revision of terminology is also inherent in clinical practice. There is a movement in nursing, for instance, to shift and standardize terminology to improve communication, patient care, and data collection.[18] For instance, "small," "moderate," or "large" amounts of bleeding have been redefined to be consistent across practice contexts. Similarly, in HPE, scholars can carefully examine terminology and find innovative ways to adapt this terminology for use across a variety of practical and theoretical contexts.[18]

Principle 5: Theories Can Be Challenged

It is often thought that theories cannot be challenged because they are based on a significant body of empirical research. We argue that theory can and should be challenged, but in HPE, "theory testing" can differ from the standard, causal comparative design or experiment. The testing of a theory may involve repeated applications in various contexts to demonstrate transferability of a theoretical concept. In quantitative research, theory testing occurs through numerical outcomes with inferential statistics to demonstrate the generalizability of the conclusion. In qualitative research, the clarity and plausibility of the findings—the extent to which they create new insights and the extent to which these insights are adopted by the scientific community—are also forms of generalization. In both cases, the research outcomes constitute the "truth" only until a "better truth" is identified.

This challenge of theory testing in HPE should not deter us from conducting further research. After all, practice devoid of theory is not useful; neither is theory devoid of practice. Indeed, we suggest that HPE offers a unique opportunity to examine educational outcomes because we educate students to become professionals across well-defined health professional domains. Thus, it is easier to link performance in practice with performance during training. We argue that theories should be continually tested and revised according to the evidence in our field because what we do affects the health care of the communities that our health professionals serve.

In medicine and public health, for example, the theory that smoking has a causal relationship with lung cancer was not tested by a single causal comparative study but, rather, by a whole program of research eventually leading to consensus. In HPE, programmatic theory testing is often more useful than a single, definitive "big bang" study. Take, for example, the challenge of understanding context specificity—a vexing medical phenomenon whereby a physician sees two

patients with identical symptoms and findings (and the same underlying diagnosis) but arrives at two different diagnostic decisions.[19] We discuss this challenge next as a context in which to understand how the 5 principles of using educational theory can be applied to a complex research problem.

A Practical Example

The phenomenon of context specificity in clinical reasoning is a complex problem. Traditionally, clinical reasoning was conceived as an individual skill, affected only by the difficulty of the content of the medical case. However, something more than the medical content is driving the physician's clinical reasoning when context specificity is observed. First, we sought to empirically investigate context specificity in the simulation environment where we could control the "stimulus" to explore this phenomenon. We carefully crafted both video and live simulation cases with identical content, which differed only by the presence or absence of contextual factors (information other than the content needed to arrive at a correct diagnosis, such as the patient being a nonnative English speaker, electronic health record malfunctions, or a fatigued physician).

We struggled with what theory to apply to the phenomenon of context specificity because there were no readily apparent HPE theories that integrated the notion of reasoning as an individual ability and how it could be affected by the health professional's environment (Principle 3: You can deviate from a theory's propositions). Therefore, we started with a macro theory, situated cognition theory (SCT), to help us operationalize some of the important features and interactions in a clinical encounter and to obtain a basic understanding of what might be important in understanding context specificity (Principle 1: All theories are not created equal). We then realized that we had to adapt this theory to our specific research context, requiring multiple

discussions and experiments.[20,21] We identified examples of aspects that played a role in a physician's clinical reasoning, other than merely the medical content, needed to arrive at the diagnosis (which we termed contextual factors). Using SCT, we grappled with these matters and grouped contextual factors related to the patient, the physician, and the encounter (see Figure 1). From this macro theory, we were able to see how such contextual factors can affect a physician's clinical reasoning, leading us to identify two more micro theories to further guide the investigation: cognitive load theory (CLT) and self-regulated learning theory (SRLT) (Principle 1: All theories are not created equal; Principle 2: Multiple theories can be used in a given research study). Using CLT, we developed measures of the potential increased mental effort (i.e., cognitive load) that might be generated as a result of the contextual factors. Although CLT is a learning theory (as opposed to a theory of performance), we adapted CLT principles for learning to evaluate performance, thereby enhancing the theory to meet our needs as HPE scholars (Principle 3: You can deviate from a theory's propositions). Then, alongside CLT, we used SRLT to explore the tools that physicians might be using to manage this increased cognitive load (e.g., strategic planning, setting goals) (see Figure 2).

The integration of these 3 different theories (SCT, CLT, and SRLT) forced us to grapple with—and reconcile—some terminology mismatches, such as the definition of clinical reasoning (Principle 4: Terminology can be reconciled across theories). We had to reconcile the meaning of the term *clinical reasoning* across multiple fields and establish a defensible meaning to help address the challenge of defining *context specificity*. Applying this integrated framework, we were able to do the following: investigate the phenomenon using a series of experiments, provide a new lens through which to understand context specificity, enhance our understanding of these

theories in a clinical sitting, and test our theoretical predictions (Principle 5: Theories can be challenged).

In this article, we have identified 5 principles for using educational theory in HPE, offering illustrative examples, including a narrative of a theoretical challenge we faced in a recent program of research examining context specificity. The complex phenomena that are characteristic of HPE can pose multiple challenges for researchers. We maintain that existing theories can offer diverse perspectives to address these challenges and move the field of HPE forward.

References

1. Cervero RM, Daley BJ. The need and curricula for health professions education graduate programs. New Directions for Adult and Continuing Education. 2018;2018(157):7-16.
2. Albert M, Hodges B, Regehr, G. Research in medical education: Balancing service and science. Advances in Health Sciences Education. 2007;12(2):103-115.
3. Ostrom E. Governing the commons: The evolution of institutions of collective action. Cambridge, UK: Cambridge University Press; 1990:45-46.
4. Swanson DB, Norcini JJ, Grosso LJ. Assessment of clinical competence: Written and computer-based simulations. Assessment and Evaluation in Higher Education. 1987;12(3):220–46.
5. Torre D, Durning SJ. Social cognitive theory: Thinking and learning in social settings. In: Cleland J, Durning SJ, eds. Researching Medical Education. Chichester, UK: Wiley Blackwell; 2015;105-116.
6. Durning SJ, Artino AR. Situativity theory: A perspective on how participants and the environment can interact: AMEE Guide no. 52. Medical Teacher. 2011;33(3):188-199.
7. Davidoff F, Dixon-Woods M, Leviton L, Michie S. Demystifying theory and its use in improvement. BMJ Qual Saf. 2015;24:228-238.
8. Reeves S, Albert M, Kuper A, Hodges BD. Why use theories in qualitative research? BMJ. 2008;337:a949.
9. Battista A, Konopasky A, Yoon M. Programs of research in healthcare simulation. In Nestel D, Hui J, Kunkler K, Calhoun A, Scerbo M, eds. Healthcare simulation research: A practical guide. Basel, Switzerland: Springer International Publishing. In press.

10. Crossley J, Davies H, Humphris G, Jolly B. Generalisability: A key to unlock professional assessment. Med Ed.2002;36(10):972-978.

11. Leppink J, van Gog T, Paas F, Sweller J. Cognitive load theory: Researching and planning teaching to maximise learning. In: Cleland J, Durning SJ, eds. Researching Medical Education. Chichester, UK: Wiley Blackwell; 2015;207-218.

12. Baker L, Shing LK, Wright S, Mylopoulos M, Kulasegaram K, Ng S. Aligning and applying the paradigms and practices of education. Acad Med. 2019;94(7):1060.

13. Greene JC. Mixed methods in social inquiry. Vol. 9. San Francisco, CA: John Wiley & Sons; 2007.

14. Maggio LA, Cate OT, Irby DM, O'Brien BC. Designing evidence-based medicine training to optimize the transfer of skills from the classroom to clinical practice: Applying the four component instructional design model. Acad Med. 2015;90(11):1457-1461.

15. Govaerts MJB, Schuwirth LWT, Van der Vleuten CPM, et al. Workplace-based assessment: Effects of rater expertise. Advances in Health Sciences Education. 2011;16(2):151-65.

16. Govaerts MJB, Van de Wiel MWJ, Schuwirth LWT, Van der Vleuten CPM, Muijtjens AMM. Workplace-based assessment: Raters' performance theories and constructs. Advances in Health Sciences Education.2013;18(3):375-396.

17. Ryan RM, Deci EL. Self-regulation and the problem of human autonomy: Does psychology need choice, self-determination, and will? Journal of Personality. 2006;74(6): 1557-1586.

18. Rutherford M. Standardized nursing language: What does it mean for nursing practice? OJIN. 2008;13(1):243-250.

19. Durning SJ, Artino AR, Boulet JR, Dorrance K, van der Vleuten C, Schuwirth L. The impact of selected contextual factors on experts' clinical reasoning performance (does context impact clinical reasoning performance in experts?). Advances in Health Sciences Education. 2012;17(1): 65-79.

20. McBee E, Ratcliffe T, Picho K, et al. Contextual factors and clinical reasoning: Differences in diagnostic and therapeutic reasoning in board certified versus resident physicians. BMC Med Educ. 2017;17(1):211.

21. Konopasky A, Ramani D, Ohmer M, et al. It totally possibly could be: How a group of military physicians reflect on their clinical reasoning in the presence of contextual factors. [Unpublished, 2019].

Figure Legends

Figure 1

Contextual factors in clinical reasoning from a situated cognition perspective.

Figure 2

Contextual factors in clinical reasoning from a multitheoretical perspective.

Figure 1

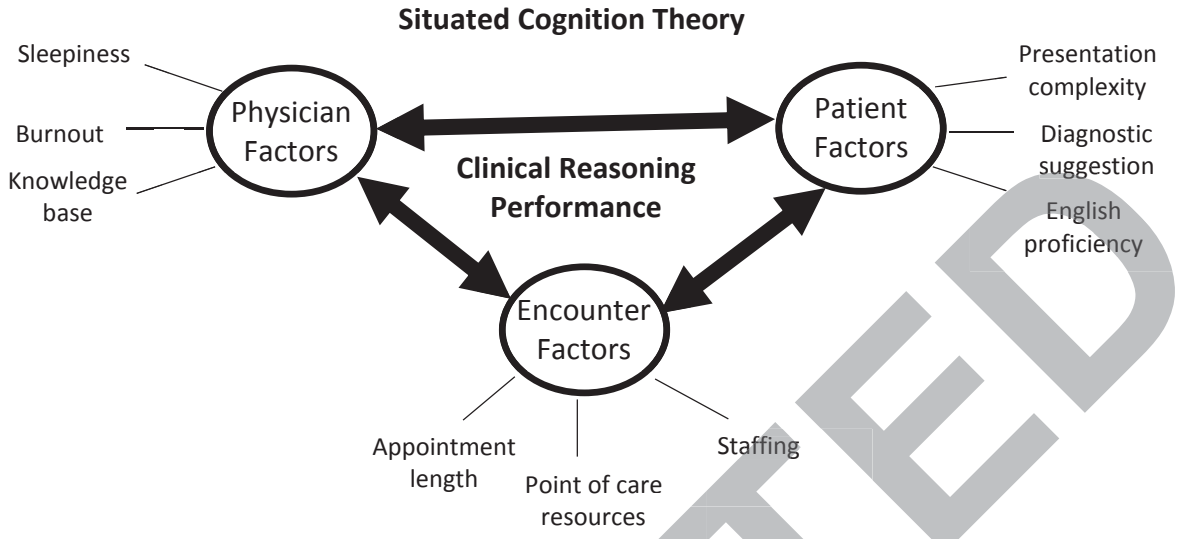
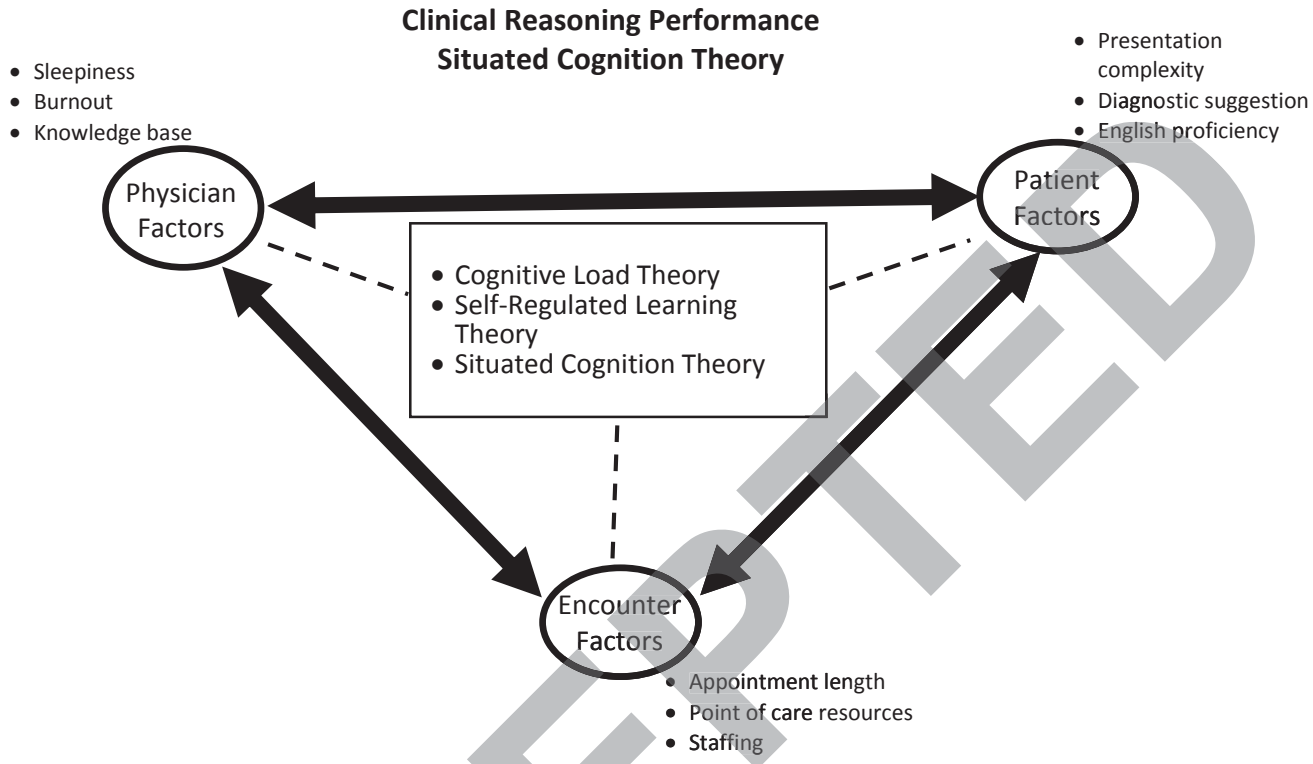


Figure 2



First-year medical students' calibration bias and accuracy across clinical reasoning activities: An initial investigation

Timothy J. Cleary, PhD,[1] Abigail Konopasky, PhD,[2] Jeffrey S. La Rochelle,[3] MPH, MD, Brian E. Neubauer, MD,[4] Steven J. Durning, MD, PhD,[2] & Anthony R. Artino, Jr., PhD[2]

Abstract

This study examined the metacognitive judgments of 157 first-year medical students as they engaged in a virtual-patient simulation activity targeting clinical reasoning practices. Examining two key subtasks of a patient encounter, history (Hx) and physical exam (PE), the authors assessed the level of variation in students' behavioral performance (i.e., effectiveness and efficiency) and judgments of performance (i.e., calibration bias and accuracy) across the two subtasks. Paired *t*-tests revealed a consistent pattern: the Hx subtask was more challenging than the PE subtask. Specifically, even though students performed worse on the Hx subtask than PE, they were less accurate in their performance judgments, with almost all participants overestimating their performance. Correlation analyses revealed that the participants' overall level of accuracy in self-judgments was fairly stable across the Hx and PE subtasks. Implications and areas for future research are discussed and analyzed.

Key words: Clinical reasoning; Metacognition; Self-assessment; Calibration; Microanalytic assessment; Self-regulated learning

Why Health Professions Education Needs Functional Linguistics:

The Power of Grammatical Categories

Abigail Konopasky, PhD, Divya Ramani, MS, Megan Ohmer, BS, Steven J. Durning, MD, PhD,

Anthony R. Artino, PhD, Alexis Battista, PhD

Graduate Programs in Health Professions Education

Uniformed Services University

Corresponding Author:

Abigail Konopasky, PhD

Uniformed Services University

Department of Medicine

Graduate Programs in Health Professions Education

4301 Jones Bridge Road

Bethesda, MD 20814

202-650-7490

abigail.konopasky.ctr@usuhs.edu

Disclaimer: The views expressed herein are those of the authors and do not necessarily reflect the official policy or position of the U.S. Navy, the Uniformed Services University of the Health Sciences, the Department of Defense, or the U.S. Government.

Acknowledgments: None

Funding Source: This paper was supported in part by a grant from the JPC - 1, CDMRP - Congressionally Directed Medical Research Program (# NH83382416).

Conflicts of Interest: None

Ethical Approval: The data reported in this paper were gathered under a research protocol approved by the Uniformed Services University Institutional Review Board.

Individual Contributions:

A Konopasky and A Battista worked together to conceive of the idea for this paper based on a joint idea.

S Durning and A Artino worked together with the previous two authors to design the structure for the paper and to draft it.

A Konopasky, D Ramani, and M Ohmer worked together to analyze and interpret the data for the portion of the paper based on the think-aloud reflections.

All authors contributed to the intellectual content of the paper through the various revision stages.

All authors have approved of the final version of this paper and agree to be accountable for ensuring that the accuracy and integrity of any part of the work is appropriately investigated and resolved should questions arise.

Abstract

Language is one of the primary modalities for teaching and learning in the health professions, from more formal teaching relationships in school to the guided practice of trainees through the continuing education and deliberate practice of lifelong learning. Yet linguistic analysis, with the possible exception of discourse analysis, has not become a core methodological tool in the field of health professions education (HPE). The purpose of this paper is to argue for more widespread adoption of one particular approach to linguistics, one that examines less *what* learners and instructors say and more *how* they say it: functional linguistics. This approach theorizes and structures the functions of language, regularly focusing attention on "forgettable" words like *I, but*, or *was*. Drawing on a rich body of literature in linguistics, psychology, the learning sciences, and some early work in HPE we demonstrate how functional linguistic tools can be applied to better understand learners' and instructors' beliefs, reasoning processes, values, and emotions. A brief qualitative analysis of one tool—the generic use of "you" to mean "one" or "anyone"—demonstrates how functional linguistics can offer insight into physicans' bids for credibility and empathy as they think aloud about their clinical reasoning. Finally, we offer suggestions for how functional linguistic tools might address questions and gaps in four active research areas in HPE: reflection, emotion and reasoning, learning in simulated contexts, and self-regulated learning. We argue that the words learners, instructors, and practitioners in the health professions use as they move through undergraduate and graduate training into practice can offer clues for researchers, instructors, and colleagues to better support them.

Meaning making through language is a powerful process that not only expresses reality, but creates it, constitutes it: "Language is the critical link between the created present and the uncreated future, affording leaders of medical schools with an underused opportunity to transform academic medicine."[1] Not only is this true for leadership contexts, but the same can be said for instructional contexts: language is one of the primary modalities for teaching and learning in the health professions, from more formal teaching relationships in school to the guided practice of trainees through the continuing education and deliberate practice of lifelong learning. All of these practices are largely constituted *by* language and transformations in these practices generally occur *through* language. In order to better understand these practices, some researchers in the field of health professions education (HPE) have taken up tools from various branches of linguistics: discourse analysis,[2,3] conversation analysis,[4,5] pragmatics,[6] natural language processing,[7,8] and functional linguistics.[9,10] Yet linguistic analysis, with the possible exception of discourse analysis, has not become a core methodological tool in the field.

The purpose of this paper is to argue for more widespread adoption of one particular approach to linguistics, one that examines less *what* learners and instructors say and more *how* they say it: functional linguistics.[11,12] At its core, this approach structures and theorizes the *functions* of different components of language. Moreover, functional analysis does not stop at the outward-referring content words, verbs like *speak* or nouns like *linguist*. It also focuses on inward-referring function words, pronouns like *I* and *we*, verb tenses like past and present, and conjunctions like *because* and *but*. In so doing, it reveals the power of grammatical categories for understanding language users' beliefs, values, and emotions. This paper is an argument for harnessing that power in HPE. We begin by describing the functional linguistic approach and arguing for its potential power as an analytic framework. We then review several studies in HPE

that have already used some of these tools, arguing for their unique contribution. Next we offer a worked example of how functional linguistics can be used in HPE, using the generic *you* pronoun. Finally, we suggest future uses for functional linguistics in HPE.

Functional Linguistics: The Power of "Stealth Words"

Functional linguistics (often called systemic functional linguistics[11,12]) is an approach to linguistics that examines how *conceptual* structures (e.g., status, identity, emotion) are represented via the *grammatical* structures that speakers and writers choose.[13] Functional linguists maintain (among other tenets) that short, "forgettable" words (e.g., pronouns, articles, prepositions--*function words*) can actually reveal much about individuals' beliefs, values, and emotions.[14]

In research programs stretching out over the past 20 years, James Pennebaker, Arthur Graesser, Danielle McNamara, and a host of colleagues have studied these "stealth words"[14] across different contexts, finding them to be associated with numerous characteristics or states.[15–17] One series of studies, for instance, used grammatical categories like articles (e.g., *the*), causal words (e.g., *because*), and negation (e.g., *not*) to discern reliable differences across individuals based on written narratives (reflections of patients in a substance abuse unit and daily reflections for a writing class).[18] The authors found moderate correlations (r = -.28, $p < .05$; r = -.30, -.33, $p < .001$,) between some of their linguistic feature variables and achievement motivation (as measured by self-report survey). These same tools could be used in HPE with the reflective writing of health professions students, for instance, to identify those who might need motivational support.

In another set of studies, pronoun use (e.g., *I, you*) and word count were found to be related to the relative status of people writing to each other. In other words, letter writers who used more words, as well as more first-person singular *I* words, were more likely to be *low* status individuals writing to *higher* status individuals (either naturally or through experimental manipulation of status).[19] In HPE, program evaluators could use these linguistic features to examine the hierarchical structures among their instructors and learners, using the results as a guide for programmatic shifts to create greater feelings of equity.

Other scholars have closely examined the structure of utterances to discern their function.[20–25] For example, one set of studies argues that patterns of subject pronouns (e.g., *I, we*) and object pronouns (e.g., *me, us*) can indicate feelings of increased or decreased agency (defined broadly as the capacity to have an intention and produce an effect based on it[26]) in an adult learning classroom.[22,27] These patterns were used to explore the kinds of actions these learners narrated themselves taking (e.g., *studying*, *trying*) and the range of other agents they narrated as acting upon them (e.g., *teachers helping, kids picking on*), creating a rich portrait of the patterns of agency in their educational experiences. A similar approach could help HPE researchers examine the patterns of agency in physicians or physicians in training, and how these might be related to important affective outcomes, like shame, well-being, or burnout.

The studies above take advantage of relatively authentic and natural uses of language to better understand individuals' implicit, and often unconscious, beliefs, values, and emotions. They do not require individuals to stop and take a survey or fill out an assessment and, unlike studies using researcher-designed surveys and assessments, these studies are grounded fully in *participants'* language and *participants'* formulations of reality. This functional linguistic approach, then, provides an analytic framework that could exploit the existing linguistic richness

inherent in many HPE contexts, such as case reports, bedside rounds, oral assessments, and chart notes. Using these sources, scholars can examine the words and grammatical structures clinicians use so that health professions educators can better understand providers' beliefs about themselves and their patients, the values they place on different sources of evidence, and the emotions they experience as they move through different practice contexts.

Stealth Words in HPE: Communication, Assessment, and Reflection

Linguistic approaches are being used more in medicine, and to great effect. Jeff Bezemer uses conversation analysis and sociolinguistics to examine operating room discourse;[28,29] Lynn Monrouxe uses discourse and narrative analysis in interview studies of medical trainees;[3,30] Molly Carnes uses automated and qualitative linguistic approaches to better understand the influence of cultural stereotypes on gender;[31,32] Lorelei Lingard uses pragmatics and conversation analysis to examine communication in health care teams;[4,6] and Debra Roter's linguistically-based Roter Interaction Analysis System (RAIS) has been used in a number of studies to explore physician-patient communication.[33,34] In addition to this work, there are also a few studies using elements of *functional* linguistics to study other health professions learning contexts. We briefly touch on three of them here, noting the tools they use and the insights they gain (see Table 1 for a summary).

First, in an article titled "The Medical Educator, the Discourse Analyst, and the Phonetician," Woodward-Kron and colleagues narrate how they leveraged functional linguistics and discourse analysis to offer linguistically-focused feedback to international medical trainees who were not practicing in their native language.[10] For instance, they distinguished between questions focused on the patient with a pronoun (e.g., Did *you*…?) and questions focused on the

symptoms or problem with a noun (e.g., Did *the pain*…?). A small case study of two of these learners found marked communication improvement, including in patient history taking.

Second, Ruitenberg and Towle draw on the linguist and philosopher J.L. Austin's work to analyze how inter-professional health care mentees "do things with words" in their journal entries.[9,35] They use functional linguistic markers of *modality*--how speakers express beliefs about certainty, obligation, or probability. This focus on written reflection (Howe and colleagues also use functional linguistics to study reflection[36]) allowed the researchers to examine the different moments in which these learners were more and less confident. They noted increased confidence (i.e., certainty) for some learners when referring to a group as a "we" compared to themselves as "I."

Third, Shapiro and colleagues use an automated tool that draws heavily from functional linguistics, Linguistic Inquiry and Word Count (LIWC), to examine how a "point-of-view" training affected medical students' written reflections on a paper case.[37] They found that those with the training used more *I* pronouns. When integrated with thematic analysis, they found that *I* pronouns were also linked with more demonstrations of empathy, a target skill for the training.

These three applications of functional linguistics--to international medical trainees' communicative skills, inter-professional health mentees' shifting confidence, and medical students' empathy--yield unique insights. Yet, even these studies do not take full advantage of the "stealth" words lurking in HPE contexts. In the next section, we offer an example of what that might look like.

Functional Linguistics and Physician Reflection: An Application of Stealth Words

Here we analyze physicians' reflections on their reasoning throughout a case using a tool derived from functional linguistics: the generic *you* pronoun. In this example, we draw on data from a broader study of physicians of varying levels of experience (first-year interns through attendings with many years of experience), in which they were asked to either participate in a live scenario of two cases or view a video of two cases (unstable angina and diabetes mellitus[38]). Afterwards, they were asked to "think aloud" about their reasoning as they either watched the video of themselves or re-watched the video of the case. The data discussed here are pulled from these "think aloud" reflections.

Generic *you*: Establishing authority and empathy. Pronouns are particularly powerful functional tools, and, in certain contexts, can signal things like relative status, relative age, depression, and even whether or not a writer is lying.[14] English is one of many languages that uses one of its personal pronouns--second person *you*--in an *impersonal* way.[39] For instance, in reference to the diabetes case, one participant noted, "**You** just don't know, did she have pituitary injury which led to her having problems controlling her water intake and output?" This *you* does not refer explicitly to the researcher conducting the think-aloud, but to *any* physician--a generic physician--who is determining a diagnosis in this case. Through the use of generic *you*, this participant invokes a broader membership category[40] that he (the participant) belongs to, *physician-in-this-case*, establishing through that membership credibility for his declaration of uncertainty about the cause of polyuria. In other words, any physician in this case would be uncertain about this. Meanwhile, by using the second-person singular pronoun *you*, the participant simultaneously aligns himself with the listener, *you-who-is-sitting-here*, seeking empathy and a shared view about uncertainty about the polyuria cause.[40] In other words, you, my conversational partner, empathize with my uncertainty, right? (See Figure 1.) This use of the

impersonal to both generalize and seek empathy allows the speaker to save face,[41] protecting a professional identity that is potentially threatened by acknowledgment of this uncertainty.

Most of these uses of *you* to align with *physician-in-this-case* involve cognitive processes like *knowing* (and *not knowing*), *getting a sense*, *having an idea of*, *wondering*, etc. In these instances, participants are choosing to reflect on what some general *you* experiences as a physician rather than saying *I*, shifting the focus away from the self to a more general situation. For instance, in reasoning through the diabetes case, one participant states, "And then **you** start thinking about why people urinate frequently." Rather than taking direct agency for what *I* am thinking, generic *you* allows participants to attribute cognitions to a broader group: it is not just what *I* think; *any* physician in this case would think this way. One potential negative outcome of this stance could be a lack of motivation to for improvement: if the self does not have control over the clinical reasoning process, there is no reason for the self to seek to *improve* the clinical reasoning process, potentially avoiding important learning opportunities.

Physicians in our study also use generic *you* to establish membership in other categories beyond *physician-in-this-case*. For instance, one first-year resident talks about why she is cautious about sharing a potential leading diagnosis with a patient: "**You** shouldn't just spew off diagnoses, at least until you've talked to **your** attending, even if they are glaringly obvious." Here she is establishing herself as a member of a more specific category, *physician-in-training*, making an argument from that membership position for not sharing a diagnosis with a patient immediately. As with the prior example, she draws her listener (an HPE researcher) in as part of the narrative, even though that listener is neither a physician nor a physician in training. The use of the second-person pronoun positions the listener as an "everyman" who is a part of things, making a bid for empathy.[39,40]

Finally, our participants position themselves as *human/patient* to lend authority to statements like "Coughing usually gets worse after **you** stop smoking." This generalized *you* positions the speaker and listener as human beings who have predictable responses to something like smoking cessation. All three of these generic membership categories--*physician-in-this-case, physician-in-training*, and *human/patient*--are used to establish credibility while also seeking empathy from the listener.

Generic "you": A case for functional linguistics in HPE research. This brief application of generic *you* to physicians' reflections on a case demonstrates some of the benefits of close analysis using functional linguistic tools. First, the different ways generic *you* is used point to the different groups physicians place themselves in as they proceed through a case (i.e., physicians generally, residents, humans/patients). These insights into participants' group memberships and precisely when participants make use of these memberships (e.g., to offer authority to a position of uncertainty) reveal details about these physicians' senses of identity and how they shift across the course of a task like reflecting on a case. Second, because generic *you* also seeks alignment with and empathy from the listener, tracing its use can indicate when in the course of a reflection physicians might be feeling the need for support or affirmation from their conversational partner. The frequent use of this pronoun with cognitive processes like *think* and *know* indicates that physicians perhaps seek support around their own thought processes (rather than identifying a symptom, for example). Third, unlike a self-report survey inquiring about these areas (i.e., identity, need for support), this tool draws directly from participants' language and experiences, allowing HPE researchers to carefully track the specific contexts and experiences that trigger certain identity affiliations or beliefs (without unintentionally biasing participants by framing questions and responses in the *researchers'* language).

Functional Linguistics and HPE: Future Directions

The above analysis of generic *you* from these think-aloud reflection data draws from a body of functional linguistic literature, both theoretical and applied, that is largely untapped in HPE. In this section, we review several research areas in HPE, suggesting potential applications of functional linguistic tools. (See Table 2 for an overview.)

Reflection. Sandars defines reflection as "a metacognitive process that occurs before, during, and after situations with the purpose of developing greater understanding of both the self and the situation so that future actions can be informed by this understanding."[42] Due to reflection's believed link with medical error,[43] it is a growing area of research. When used appropriately, reflection can contribute to learning.[42,44–46] Yet we are only beginning to carefully evaluate and assess the *effects* of different approaches to reflective practice.[42,46,47] Since reflection is usually a relatively open-ended process, it can be difficult to capture the distinctions among reflections and determine which of those distinctions may be indicative of some kind of important cognitive, affective or social change. Functional linguistic tools can be applied to reflections of any format or length, offering potential ways to both describe and quantify reflective practice. For instance, instructors could track students' personal pronoun use over time as they reflect on developing a diagnosis for a case to see if their attentional focus shifts from the self's performance (i.e., *I*) to the patient's concerns (i.e., *she/he*) and symptoms and diagnosis (i.e, *it*).[14] Moreover, instructors could share these assessment tools with students in order to help them self-assess and consciously redirect their reflections.

Emotions and reasoning. In a recent review, Kozlowski and colleagues conclude that while clinicians' emotions can and do influence clinical reasoning, this influence is often not acknowledged in clinical reasoning research and training.[48] Much of the research, then, focuses on the cognitive components without accounting for the important role of emotions in reasoning. They argue that more research is needed in order to "provide a more thorough and intentional picture" of the ways that emotions (positive and negative) interact with reasoning processes and outcomes. There is a robust tradition of using tools from functional linguistics to investigate emotions.[15,49] One potential application to HPE is to examine the language learners use as they reason through a past case for their use of tense (i.e., past, present, future). Shifting rapidly between past and present, for instance, can indicate the writer is reflecting on a traumatic experience, while focusing only on the past or future (rather than the present) can indicate sadness.[14] This could provide a better understanding of the emotional aspect of reasoning, particularly in the presence of affective contextual factors like a frustrated patient or anxiety about one's uncertainty or confusion about the diagnosis.[50,51] This would allow researchers to develop better support for these contextual factors so that these emotional dimensions do not interfere with diagnosis and treatment. Using functional linguistics in this way supports recent calls for HPE researchers to use less intrusive measurement techniques when studying emotions.[52]

Learning in simulated contexts. Simulated contexts, which allow learners to engage in focused practice, are also commonly used for learner assessment.[53] The increasing popularity and effectiveness of simulation in healthcare has led a number of researchers to call for more careful attention to the ways that learning processes and outcomes are measured in simulated contexts.[53–55] One aspect of this is the need to better understand how different features of the simulated

environment affect learner outcomes, particularly as they relate to preparation for future learning.[55] The language data generated during a simulated encounter—the participant's utterances during simulation, post-simulation debriefings, and participants' written patient notes, to name a few—provide numerous opportunities for measuring learner outcomes with functional linguistics. For example, conjunctions like *and, but,* and *because* indicate that a speaker is connecting multiple statements together, whether bringing two ideas together with a word like *and*, drawing distinctions with a word like *but*, or making a causal claim with a word like *because*. Words like these can be indicators of cognitive complexity[15] and, as such, could help distinguish among simulation participants at varying performance levels. Functional linguistic tools like these offer a novel methodological approach to simulation, increasing the variety of available research methods and, thus, providing a different lens on simulated contexts.[54]

Self-regulated learning. Self-regulated learning (SRL)--the processes by which learners moderate their actions, experiences, and environment towards some learning goal[56]--has become an active area of research in HPE, allowing researchers to better study learning strategies across classroom and clinical environments.[57] Much of this research, however, has relied on learners' self-reports of SRL strategies and, moreover, has taken place more in experimental laboratory settings rather than the messy, "real world" settings of clinical practice.[58,59] Innovative methods for dynamically assessing the *process* of SRL are needed in order to better understand the steps learners take in these authentic clinical settings and how these steps differ across individuals and contexts.[56,58,59] Functional linguistics is well positioned to offer some of these innovative methods. For instance, modal verbs (auxiliary verbs expressing possibility or obligation) indicating a sense of obligation (e.g., *should, have to*) can indicate diminished feelings of agency[22,25,49] (e.g., "I **have to** finish the review of systems" versus "I want to/am going to finish

the review of systems"). HPE scholars could use these modals to track shifts in feelings of agency for SRL strategies either within individuals over time or between individuals across different clinical environments.

Conclusions

Drawing on a rich body of literature in linguistics, psychology, and the learning sciences, we argue here for the adoption of more functional linguistic tools in HPE. While several studies have incorporated some aspects of functional linguistics,[9,10,37] none have taken advantage of the range of tools available for better understanding beliefs, values, and emotions through language. Through our brief qualitative analysis of generic "you" in physician think-aloud reflections on a clinical case, we argued that participants used the impersonal and personal meanings of "you" in order to seek credibility and authority on the one hand and empathy and alignment on the other. These bids for credibility and empathy occurred most around cognitive processes, suggesting perhaps that those were statements for which these physicians sought broader support for their claims. Finally, we offered specific functional linguistic tools to address questions and gaps in four research areas: reflection, emotion and reasoning, learning in simulated contexts, and SRL. Tools like pronouns, tense, conjunctions, and modals, while seemingly forgettable, could aid with assessment and evaluation across these areas.

As researchers and practitioners begin to tune in more carefully to patient narratives, really listening to the emotions, beliefs, and values behind patient stories,[60–62] it is perhaps time for them to tune in to their own language as well. The words health professionals use as they move through undergraduate training, graduate training, and lifelong practice can offer clues for researchers, instructors, and colleagues to better support them. While grammatical markers like

conjunctions, prepositions, and verb tense may be "forgettable" for the person who uses them, they are, in James Pennebaker's words, "glorious language markers"[14] researchers and practitioners can use to broaden and deepen the work of HPE.

References

1. Souba C. Perspective: The language of leadership. *Acad Med*. 2010;85(10):1609-1618.
2. Chan MY. The oral case presentation: toward a performance-based rhetorical model for teaching and learning. *Med Educ Online*. 2015;20(1):28565.
3. Rees CE, Cleland JA, Dennis A, Kelly N, Mattick K, Monrouxe L V. Supervised learning events in the Foundation Programme: a UK-wide narrative interview study. *BMJ Open*. 2014;4(10):e005980.
4. Kennedy TJT, Lingard LA. Questioning competence: A discourse analysis of attending physicians' use of questions to assess trainee competence. *Acad Med*. 2007;82(10):S12-S15.
5. Roberts C, Wass V, Jones R, Sarangi S, Gillett A. A discourse analysis study of 'good'and 'poor'communication in an OSCE: a proposed new framework for teaching students. *Med Educ*. 2003;37(3):192-201.
6. Ginsburg S, van der Vleuten C, Eva KW, Lingard L. Hedging to save face: a linguistic analysis of written comments on in-training evaluation reports. *Adv Heal Sci Educ*. 2016;21(1):175-188.
7. Mata DA, Ramos MA, Kim MM, Guille C, Sen S. In their own words: an analysis of the experiences of medical interns participating in a prospective cohort study of depression. *Acad Med J Assoc Am Med Coll*. 2016;91(9):1244.

8. Zhang R, Pakhomov S, Gladding S, Aylward M, Borman-Shoap E, Melton GB. Automated assessment of medical training evaluation text. In: *AMIA Annual Symposium Proceedings*. Vol 2012. American Medical Informatics Association; 2012:1459.
9. Ruitenberg CW, Towle A. "How to do things with words" in health professions education. *Adv Heal Sci Educ*. 2015;20(4):857-872.
10. Woodward-Kron R, Stevens M, Flynn E. The medical educator, the discourse analyst, and the phonetician: A collaborative feedback methodology for clinical communication. *Acad Med*. 2011;86(5):565-570.
11. Halliday MAK, Matthiessen C, Halliday M. *An Introduction to Functional Grammar*. Routledge; 2014.
12. Martin JR, Rose D. *Working with Discourse: Meaning beyond the Clause*. Bloomsbury Publishing; 2003.
13. Lakoff G, Johnson M. *Metaphors We Live By*. University of Chicago press; 2008.
14. Pennebaker JW. The Secret Life of Pronouns: What Our Words Say About Us (NY: Bloomsbury). 2011.
15. Tausczik YR, Pennebaker JW. The psychological meaning of words: LIWC and computerized text analysis methods. *J Lang Soc Psychol*. 2010;29(1):24-54.
16. McNamara DS, Louwerse MM, McCarthy PM, Graesser AC. Coh-Metrix: Capturing linguistic features of cohesion. *Discourse Process*. 2010;47(4):292-330.
17. Graesser AC, McNamara DS, Louwerse MM, Cai Z. Coh-Metrix: Analysis of text on cohesion and language. *Behav Res methods, instruments, Comput*. 2004;36(2):193-202.
18. Pennebaker JW, King LA. Linguistic styles: Language use as an individual difference. *J Pers Soc Psychol*. 1999;77(6):1296.

19. Kacewicz E, Pennebaker JW, Davis M, Jeon M, Graesser AC. Pronoun use reflects standings in social hierarchies. *J Lang Soc Psychol*. 2014;33(2):125-143.

20. Steinkuehler CA. Massively multiplayer online video gaming as participation in a discourse. *Mind, Cult Act*. 2006;13(1):38-52.

21. Kennedy-Clark S, Thompson K. Using multimodal discourse analysis to identify patterns of problem solving processes in a computer-supported collaborative environment. In: *Innovative Methods and Technologies for Electronic Discourse Analysis*. IGI Global; 2014:94-118.

22. Konopasky AW, Sheridan KM. Towards a Diagnostic Toolkit for the Language of Agency. *Mind, Cult Act*. 2016;23(2). doi:10.1080/10749039.2015.1128952

23. Quigley J. *The Grammar of Autobiography: A Developmental Account*. Psychology Press; 2000.

24. Rogers R. Storied selves: A critical discourse analysis of adult learners' literate lives. *Read Res Q*. 2004;39(3):272-305.

25. Rymes B. *Conversational Borderlands: Language and Identity in an Alternative Urban High School*. Teachers College Press; 2001.

26. Kaptelinin V, Nardi BA. *Acting with Technology: Activity Theory and Interaction Design*. MIT press; 2006.

27. Konopasky AW. A Linguistic Approach to Agency: A Mixed Methods Study of Adult Secondary Learners. 2016.

28. Bezemer J, Murtagh G, Cope A, Kneebone R. Surgical decision making in a teaching hospital: a linguistic analysis. *ANZ J Surg*. 2016;86(10):751-755.

29. Bezemer J, Cope A, Korkiakangas T, et al. Microanalysis of video from the operating

room: an underused approach to patient safety research. *BMJ Qual Saf*. 2017;26(7):583-587.

30. Dennis AA, Foy MJ, Monrouxe L V, Rees CE. Exploring trainer and trainee emotional talk in narratives about workplace-based feedback processes. *Adv Heal Sci Educ*. 2018;23(1):75-93.

31. Kaatz A, Magua MW, Zimmerman DR, Carnes M. A quantitative linguistic analysis of National Institutes of Health R01 application critiques from investigators at one institution. *Acad Med J Assoc Am Med Coll*. 2015;90(1):69.

32. Isaac C, Chertoff J, Lee B, Carnes M. Do students' and authors' genders affect evaluations? A linguistic analysis of medical student performance evaluations. *Acad Med*. 2011;86(1):59.

33. Roter DL, Larson SM, Beach MC, Cooper LA. Interactive and evaluative correlates of dialogue sequence: a simulation study applying the RIAS to turn taking structures. *Patient Educ Couns*. 2008;71(1):26-33.

34. October TW, Dizon ZB, Roter DL. Is it my turn to speak? An analysis of the dialogue in the family-physician intensive care unit conference. *Patient Educ Couns*. 2018;101(4):647-652.

35. Austin JL. *How to Do Things with Words*. Vol 88. Oxford university press; 1975.

36. Howe A, Barrett A, Leinster S. How medical students demonstrate their professionalism when reflecting on experience. *Med Educ*. 2009;43(10):942-951.

37. Shapiro J, Rucker L, Boker J, Lie D. Point-of-view writing: a method for increasing medical students' empathy, identification and expression of emotion, and insight. *Educ Heal*. 2006;19(1):96-105.

38. Alexis Battista Divya Ramani, Megan Ohmer, Jeffrey Mikita, Anna Howle, Sarah Krajnik, Dario Torre, Steven J Durning AK. Clinical Reasoning in the Primary Care Setting: Two Scenario-Based Simulations for Residents and Attendings. *MedEdPORTAL Publ*.

39. Kitagawa C, Lehrer A. Impersonal uses of personal pronouns. *J Pragmat*. 1990;14(5):739-759.

40. Stirling L, Manderson L. About you: Empathy, objectivity and authority. *J Pragmat*. 2011;43(6):1581-1602.

41. Levinson P, Brown P, Levinson SC, Levinson SC. *Politeness: Some Universals in Language Usage*. Vol 4. Cambridge university press; 1987.

42. Sandars J. The use of reflection in medical education: AMEE Guide No. 44. *Med Teach*. 2009;31(8):685-695.

43. Mamede S, Splinter TAW, van Gog T, Rikers RMJP, Schmidt HG. Exploring the role of salient distracting clinical features in the emergence of diagnostic errors and the mechanisms through which reflection counteracts mistakes. *BMJ Qual Saf*. 2012;21(4):295-300.

44. Mamede S, Schmidt HG. Reflection in medical diagnosis: a literature review. *Heal Prof Educ*. 2017;3(1):15-25.

45. Ng SL, Kinsella EA, Friesen F, Hodges B. Reclaiming a theoretical orientation to reflection in medical education research: a critical narrative review. *Med Educ*. 2015;49(5):461-475.

46. Winkel AF, Yingling S, Jones A-A, Nicholson J. Reflection as a learning tool in graduate medical education: A systematic review. *J Grad Med Educ*. 2017;9(4):430-439.

47. Fragkos K. Reflective practice in healthcare education: an umbrella review. *Educ Sci*. 2016;6(3):27.

48. Kozlowski D, Hutchinson M, Hurley J, Rowley J, Sutherland J. The role of emotion in clinical decision making: an integrative literature review. *BMC Med Educ*. 2017;17(1):255.

49. Capps L, Ochs E. *Constructing Panic*. Harvard University Press; 1995.

50. McBee E, Ratcliffe T, Picho K, et al. Consequences of contextual factors on clinical reasoning in resident physicians. *Adv Heal Sci Educ*. 2015;20(5):1225-1236.

51. Konopasky AW, Artino AR, Battista A, et al. *Understanding Context Specificity: The Effect of Contextual Factors on Clinical Reasoning*. Bethesda, MD

52. Artino Jr AR, Naismith LM. 'But how do you really feel?'Measuring emotions in medical education research. *Med Educ*. 2015;49(2):140-142.

53. Cook DA, Brydges R, Zendejas B, Hamstra SJ, Hatala R. Technology-enhanced simulation to assess health professionals: a systematic review of validity evidence, research methods, and reporting quality. *Acad Med*. 2013;88(6):872-883.

54. Dieckmann P, Phero JC, Issenberg SB, Kardong-Edgren S, Østergaard D, Ringsted C. The first Research Consensus Summit of the Society for Simulation in Healthcare: conduction and a synthesis of the results. *Simul Healthc*. 2011;6(7):S1-S9.

55. Issenberg SB, Ringsted C, Østergaard D, Dieckmann P. Setting a research agenda for simulation-based healthcare education: a synthesis of the outcome from an Utstein style meeting. *Simul Healthc*. 2011;6(3):155-167.

56. Cleary TJ, Durning SJ, Artino AR. Microanalytic assessment of self-regulated learning during clinical reasoning tasks: recent developments and next steps. *Acad Med*.

2016;91(11):1516-1521.

57. van Houten-Schat MA, Berkhout JJ, van Dijk N, Endedijk MD, Jaarsma ADC, Diemers AD. Self-regulated learning in the clinical context: a systematic review. *Med Educ*. 2018:1-8. doi:10.1111/medu.13615

58. Artino AR, Brydges R, Gruppen LD. Self-regulated learning in healthcare profession education: theoretical perspectives and research methods. *Res Med Educ*. 2015:155-166.

59. Cleary TJ, Durning S, Gruppen L, Hemmer P, Artino A. Self-regulated learning in medical education. *Oxford Textb Med Educ*. 2013:465-477.

60. Ofri D. *What Patients Say, What Doctors Hear*. Beacon Press; 2017.

61. Charon R. Narrative medicine: a model for empathy, reflection, profession, and trust. *Jama*. 2001;286(15):1897-1902.

62. Charon R, DasGupta S, Hermann N, Marcus ER, Spiegel M. *The Principles and Practice of Narrative Medicine*. Oxford University Press; 2016.

Sequence Matters: Patterns in Task-Based Clinical Reasoning

Short title: Sequence Matters

Michael Soh, Abigail Konopasky, Steven J Durning, Divya Ramani, Elexis McBee, Temple Ratcliffe, Jerusalem Merkebu

Corresponding author:

Michael Soh, Assistant Professor, 4301 Jones Bridge Road, Bethesda, MD 20814

Keywords:

task-based clinical reasoning, situated cognition, situativity, social cognitive theory

Word count: 3388

Number of tables and figures: 7

Supplemental material?: No

Introduction

Background

Clinical reasoning is a complex phenomenon that involves taking a history, performing a physical exam, ordering and interpreting laboratory and/or radiographic tests (at times) as well as designing a management plan that is appropriate for a patient's circumstances and preferences (1-3). It is comprised of processes that allow the clinician to properly diagnose and manage an illness, analogously, requiring wisdom, insight, and experience (4). Deconstructing this inherent complexity somewhat, Juma and Goldschmidt delineated 26 tasks that physicians may implement when encountering patients (1). Through stages of framing, diagnosis, management and reflection, they identified that physicians assess priorities, differentiate most likely diagnosis, and establish management plans (1). Similarly, McBee and colleagues studied diagnostic and therapeutic reasoning comprised of the aforementioned framework using carefully constructed videotapes (5). McBee et al. highlight the role of context and expert performance in driving variability across clinical encounters and how physicians employ a variety of clinical reasoning tasks (5). Additionally, their study findings suggested that resident physicians used tasks in a varied and non-sequential manner (5). Both of these sets of authors speculate about patterns that might emerge in these tasks to help us better understand the process of clinical reasoning; in this study, we begin this process.

Situativity and Context Specificity

Our investigation was informed by situativity theory, and in particular situated cognition (6). Situated cognition is grounded in the notion that thinking (in this case clinical reasoning) is *situated* (or located) in experience, dividing the clinical experience

into physician, patient, and environmental facets (factors) and their interactions (7). Clinical reasoning is believed to emerge from these various interactions and thus, modification of or differences in these situation specific elements would be expected to impact clinical reasoning and potentially not only a physician's diagnostic and management decisions but also the path that they take to arrive at these decisions.

Evidence for this situated cognition approach includes the finding of context specificity (8). Context specificity is the phenomenon whereby a physician arrives at two different diagnoses for two patients with the same symptoms and findings as well as underlying diagnosis, but different situations. Durning et al. explored the interactions among the physician, patient, and environment in their study of contextual factors (alterations of context such as a patient suggesting a diagnosis, the electronic medical record improperly functioning or the physician being sleep deprived—patient, environment and physician contextual factors, respectively) and how they impact diagnostic reasoning of board-certified internists (7). They found that experts' performance was impacted by contextual factors and this observed impact was consistent with the tenets of situated cognition and cognitive load theories. Cognitive load theory pertains to our limited cognitive architecture and how we can only hold or process a limited number of pieces of information in our short term (or working) memory. High cognitive load is believed to be a potential mechanism underpinning context specificity.

Here, we investigate the presence of emerging patterns – both sequential (from case beginning to end across all participants) and comparative (comparing the aforementioned sequences in cases with and without contextual factors)– that we

believe could yield another perspective into context specificity and situated cognition, revealing how clinical reasoning tasks can unfold for physicians. First, we sought to determine if a discernable pattern appears to emerge in physicians' clinical reasoning. Second, we explored whether specific contextual factors influenced the observed patterns. And finally, we sought to delineate whether these situation-specific (contextual factors present or not) patterns provide additional evidence of the impact of contextual factors on clinical reasoning performance.

To our knowledge this is the first attempt to describe the patterns extant in clinical reasoning tasks and how they might be altered by contextual factors. Consistent with situated cognition, we hypothesize that this is not a linear process, rather one that is emergent as clinical reasoning is iterative in nature. More specifically, we predicted that the sequence of task categories would unfold in a way similar to a typical clinical encounter. In other words, broadly, the clinical reasoning process begins with taking a patient's history (framing), performing a physical and generating a diagnosis (diagnosis), developing a management plan (management), and finally, reflecting on their overall reasoning process (reflection). Within this broad pattern, we predicted smaller iterative cycles of some of these task types. Such findings could help inform our understanding of clinical reasoning and context specificity, provide additional evidence for situated cognition as an appropriate theory for exploring clinical reasoning, and the emergence of a lucid pattern may provide a useful framework for teaching and assessing clinical reasoning processes.

Materials/Methods

Sequence Visualization of Task Categories

The sample included for this analysis was derived from a larger study focused on clinical reasoning performance (9). Physicians either viewed a videotape or participated in a standardized patient encounter and then they completed a post-encounter form (PEF) (10). Immediately following PEF completion, they were asked to think aloud about how they reached their diagnosis and treatment plan while either re-watching the videotape or watching a video of the standardized patient encounter. Physicians viewed two videotapes or participated in two standardized patient encounters. In short, 60 think-aloud transcripts from 30 physicians who participated in two separate cases – one with a contextual factor and one without – were coded for 26 clinical reasoning tasks (1). These tasks were organized temporally, i.e. by order along their think-aloud process. For the purpose of this paper, we refer to the order in which tasks were demonstrated during a think-aloud as "steps". For example, a participant may exhibit task 3 as their first step, task 8 as their fourth step, and task 24 as their final step.

Frequencies of each of the 26 tasks were aggregated and categorized by *framing* the encounter (tasks 1-3), *diagnosis* (tasks 4-11), *management* (tasks 12-24), and *reflection* (25 and 26). These categorical counts were visualized using stacked area charting via Microsoft Excel to better understand the emerging, sequential nature of clinical reasoning tasks. Additionally, based on the frequencies, percentages were also calculated in order to numerically visualize the task category sequences.

'Crests' in Clinical Reasoning Task Categories

In addition to visualizing said data, based on percentages of task categories, we sought to identify ‘crests’ in the demonstration of clinical reasoning tasks. For the purposes of this study, the highest ‘crest’, or surge in the percentage of tasks in a particular category, was identified in each task category. For example, the largest ‘crest’ constitutes the highest average percentage of tasks during the think-alouds. We employed this crest visualization technique in an effort to better understand when particular task categories peaked in usage by our study participants. We highlight each ‘crest’ in our aggregate sample and compare ‘crests’ between cases with and without contextual factors and participant clinical experience.

Additionally, because each think-aloud accounts for one task count per step, a range of 0-60 task counts (representing 0-60 think-alouds) could be observed for each step. As such, any steps that amounted to less than 11 task counts (or 11 think-alouds) were removed from the dataset. For the purposes of our study, this ensured that all four task categories were consistently and adequately represented and prevented skewed percentages as a result of extended think-alouds that tapered off into a repetitive, single task count. This resulted in different ranges, but allowed for a cleaner and more accurate representation of the ‘cresting’ within the task categories.

Results

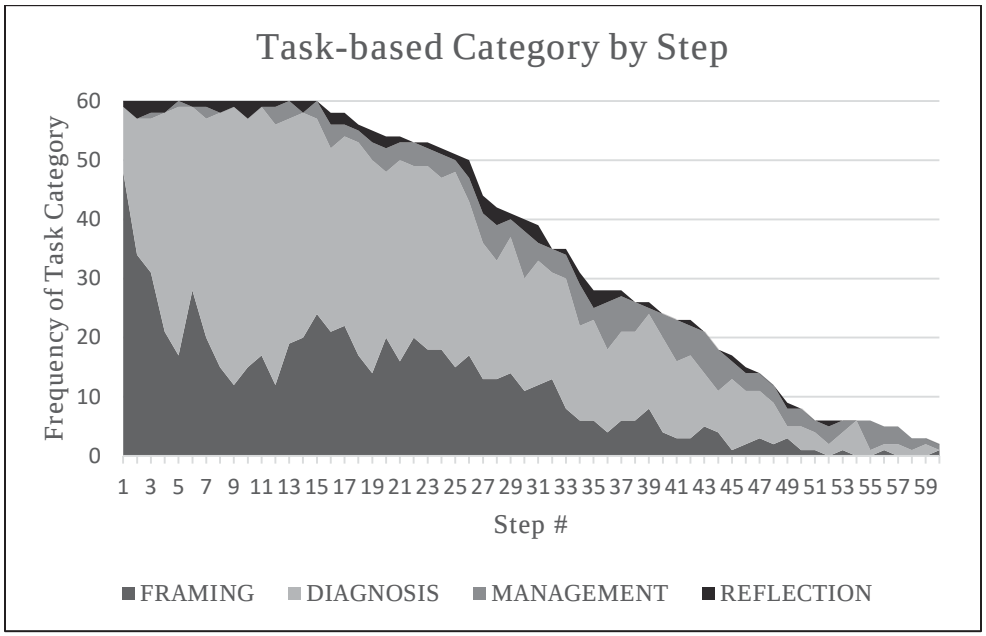
The range of steps per participant was 15-85 and the average number of steps for all physicians was 36.58 (SD=13.39). In other words, one physician required 15 steps in order to complete their think-aloud while another required 85, regardless of contextual

factor presence. This range is essential because it suggests the importance of all the steps required to arrive at a diagnosis.

General Sequence and Pattern

Our first visualization represents all 60 transcripts (or 30 physicians), i.e. think-alouds from all physicians regardless of contextual factor presence. This allowed us to view the temporal nature of clinical reasoning tasks as it unfolded during the think-alouds, in aggregate.

Figure 1. Task-based Category by Step (n=60)



During the first fifteen steps, regardless of contextual factor presence or not, all participants are engaged in a clinical reasoning task. At step 15, our first participant concludes their think-aloud and as a result, our total task count begins its decline. At step 1, 80% of participants are engaged in a framing task, 18% in a diagnosis tasks, 0%

in a management task, and 2% in a reflection task. As the think-aloud progresses into steps 8-12, framing tasks decline sharply (~23%) and diagnosis tasks increase (~75%). Moving further along the think-aloud process, we see management tasks begin to pick up at step 36 and continue until the end of the think-aloud. Reflection seems to occur throughout the think-aloud process; though, the two reflection tasks seem to peak at step 35.

In order to further illustrate the sequential nature of the task-based categories, the following table showcases the percentage calculations as bands of intensity. In other words, the darker the band, the higher the frequency of a given task category (framing, diagnosis, management or reflection). The steps within a 'crest' are highlighted in red.

Table 1. 'Crests' in Clinical Reasoning Task Categories (n=60)

Step	Framing	Diagnosis	Management	Reflection
1	80%	18%	0%	2%
2	57%	38%	0%	5%
3	52%	43%	2%	3%
4	35%	62%	0%	3%
5	28%	70%	2%	0%
6	47%	52%	0%	2%
7	33%	62%	3%	2%
8	25%	72%	0%	3%
9	20%	78%	0%	2%

10	25%	70%	0%	5%
11	28%	70%	0%	2%
12	20%	73%	5%	2%
13	32%	63%	5%	0%
14	33%	63%	0%	3%
15	40%	55%	5%	0%
16	36%	53%	7%	3%
17	38%	55%	3%	3%
18	30%	64%	4%	2%
19	25%	65%	5%	4%
20	37%	52%	7%	4%
21	30%	63%	6%	2%
22	38%	55%	8%	0%
23	34%	58%	6%	2%
24	35%	56%	8%	2%
25	29%	65%	4%	2%
26	34%	52%	8%	6%
27	30%	52%	11%	7%
28	31%	48%	14%	7%
29	34%	56%	7%	2%
30	28%	48%	20%	5%
31	31%	54%	8%	8%
32	37%	51%	11%	0%

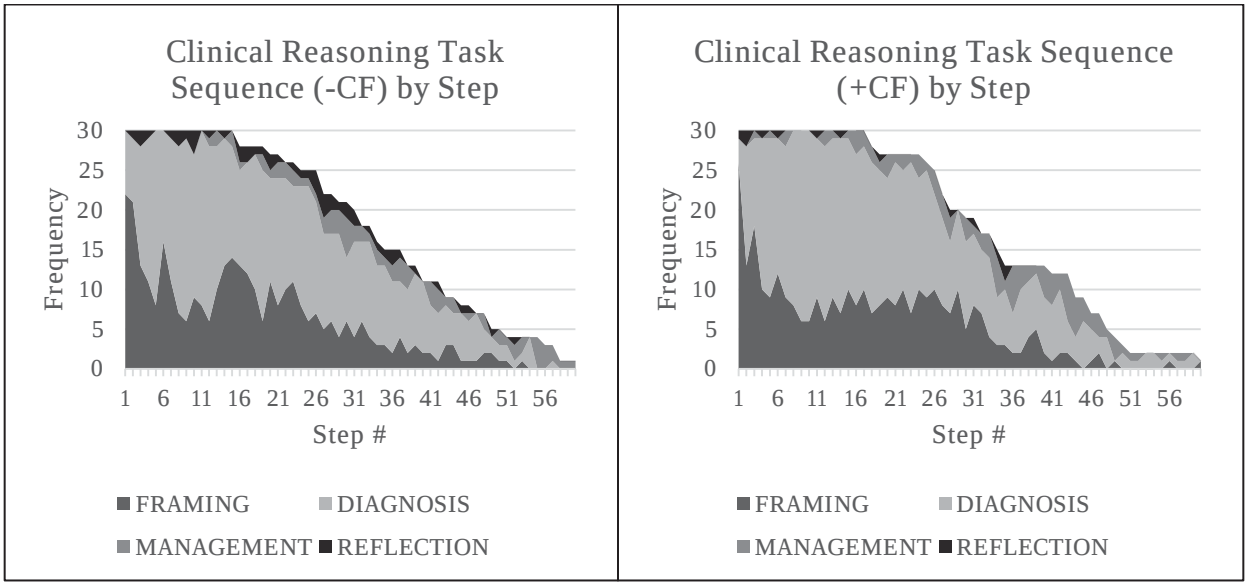
33	23%	63%	11%	3%
34	19%	52%	23%	6%
35	21%	61%	7%	11%
36	14%	50%	29%	7%
37	21%	54%	21%	4%
38	23%	58%	19%	0%
39	31%	62%	4%	4%
40	17%	67%	17%	0%
41	13%	57%	30%	0%
42	13%	61%	22%	4%
43	24%	43%	33%	0%
44	22%	39%	39%	0%
45	6%	71%	18%	6%
46	13%	60%	20%	7%
47	21%	57%	21%	0%
48	17%	58%	25%	0%

As Table 1 indicates, framing tasks peak during steps 1-3, diagnosis tasks peak during steps 8-12, management tasks peak during 36-44, and reflection tasks peak during steps 27-36.

Contextual Factor Patterns

The difference between cases with and without contextual factors is shown in Figure 2. In the absence of contextual factors the lines demonstrate larger increases and decreases of various task categories, primarily when framing the encounter. This variability in amplitude continues throughout the think aloud but is much more prominent through step 25. Conversely, in the presence of contextual factors, the lines appear less varied. This trend is consistent across the four clinical reasoning task categories. The visual representations depicted below capture this phenomenon.

Figure 2. Task-based Category by Step, without and with a contextual factor (n=30)



In the absence of a contextual factor, not only does there seem to be more variability while framing the encounter, but framing tasks also seem to be occurring at a higher rate. Also, in the absence of a contextual factor, management tasks do not occur until much later in the think-aloud process (step 12 versus step 3 in the presence of a contextual factor). And finally, as documented by the thicker black band across the top

of both charts, reflection seems to occur most consistently and at a much higher rate in the absence of a contextual factor.

Contextual Factors and Experience

The differences in patterns in the absence and presence of contextual factors seem to resemble the differences in patterns between more and less experienced physicians (over a decade [practicing physicians] compared to less than a decade [residents]). The fluid up and down movements are present in the reasoning tasks displayed by experienced physicians in the absence of contextual factors. The figures and tables below show how experienced physicians become more like less experienced physicians in the presence of contextual factors, which serve as cognitive "distractors".

Figure 3. Task-Based Category by Step, No Contextual Factor (n=30) vs Experienced Physicians (n=18)

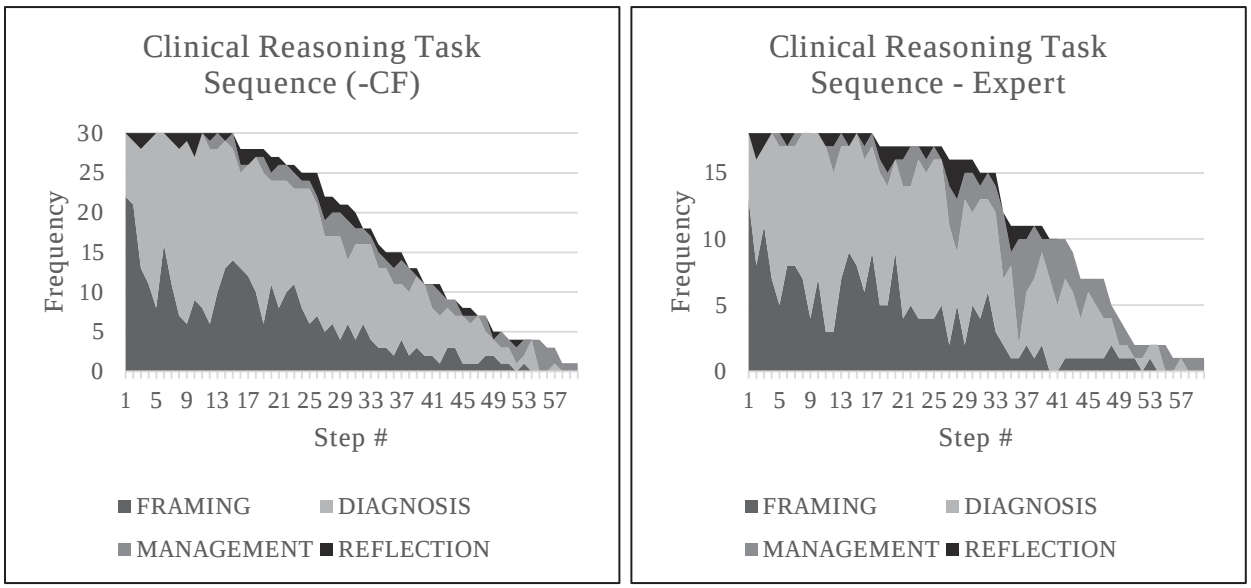


Table 2 further delineates the relationship between clinical reasoning task sequencing in the absence of a contextual factor and that of an experienced physician. The 'crests' – highlighted in red and also indicated by darker shaded bands in Tables 2 and 3 – are occurring at nearly identical sequence points for experienced physicians (regardless if a contextual factor was present) and physicians reasoning in the absence of a contextual factor. In other words, we are seeing the highest 'crest' forming at steps 1-3 (framing), 9-11 (diagnosis), 32-38 (management), and 26-28 (reflection) in both think-aloud groups.

Table 2. 'Crests' in Clinical Reasoning Task Categories (all -CF think-alouds, n=30 vs. expert think-alouds (+CF and -CF, n=18)

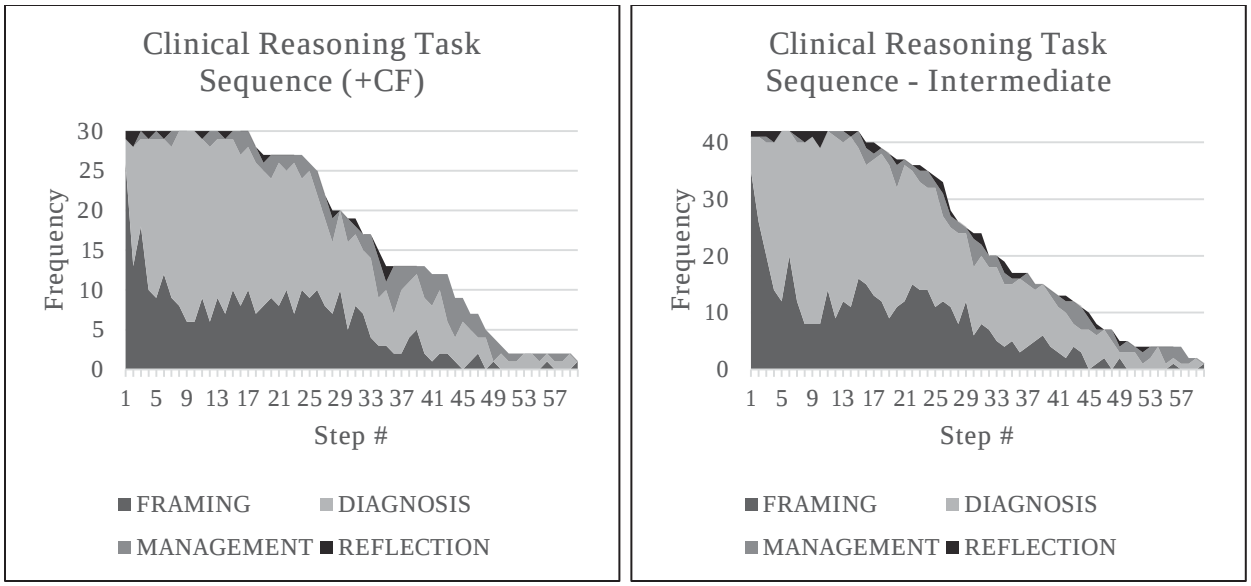
	No Contextual Factor				Expert Physicians			
step	F	D	M	R	F	D	M	R
1	73%	27%	0%	0%	72%	28%	0%	0%
2	70%	27%	0%	3%	44%	44%	0%	11%
3	43%	50%	0%	7%	61%	33%	0%	6%
4	37%	60%	0%	3%	39%	61%	0%	0%
5	27%	73%	0%	0%	28%	67%	6%	0%
6	53%	47%	0%	0%	44%	50%	0%	6%
7	37%	60%	0%	3%	44%	50%	6%	0%
8	23%	70%	0%	7%	39%	61%	0%	0%
9	20%	77%	0%	3%	22%	78%	0%	0%
10	30%	60%	0%	10%	39%	61%	0%	0%

11	27%	73%	0%	0%	17%	78%	0%	6%
12	20%	73%	3%	3%	17%	67%	11%	6%
13	33%	60%	7%	0%	39%	56%	6%	0%
14	43%	53%	0%	3%	50%	44%	0%	6%
15	47%	47%	7%	0%	44%	56%	0%	0%
16	46%	43%	4%	7%	33%	56%	6%	6%
17	43%	50%	0%	7%	50%	44%	6%	0%
18	36%	61%	0%	4%	29%	59%	6%	6%
19	21%	68%	7%	4%	29%	53%	6%	12%
20	41%	48%	4%	7%	53%	41%	0%	6%
21	30%	59%	7%	4%	24%	59%	12%	6%
22	38%	54%	8%	0%	29%	53%	18%	0%
23	42%	46%	8%	4%	24%	71%	6%	0%
24	32%	60%	4%	4%	24%	65%	6%	6%
25	24%	68%	4%	4%	24%	71%	6%	0%
26	28%	56%	4%	12%	29%	65%	0%	6%
27	23%	55%	9%	14%	13%	56%	19%	13%
28	27%	50%	14%	9%	31%	25%	25%	19%
29	19%	62%	14%	5%	13%	69%	13%	6%
30	29%	38%	24%	10%	31%	44%	19%	6%
31	20%	60%	10%	10%	27%	60%	7%	7%
32	33%	56%	11%	0%	40%	47%	13%	0%

33	22%	67%	6%	6%	20%	60%	13%	7%
34	19%	63%	13%	6%	17%	42%	42%	0%
35	20%	67%	7%	7%				
36	13%	60%	13%	13%				
37	27%	47%	20%	7%				
38	15%	62%	23%	0%				
39	23%	69%	0%	8%				

Figure 4 offers a different glimpse into the patterns and sequences of intermediate physicians (in the presence and absence of a contextual factor) and all physicians performing in the presence of a contextual factor. The light gray band representing diagnosing tasks is much thicker for both groups and management and reflection tasks are exercised much less frequently (as noted by the thinner gray and black bands).

Figure 4. Task-based Category by Step, Contextual Factor (n=30) vs Intermediate Physicians (n=42)



As shown in Table 3, the 'crests' observed in contextual factor think-alouds and intermediate physician think-alouds are synchronous. The highest 'crest' observed in framing occurs during steps 1-3, in diagnosis during steps 8-10, in management during steps 41-43, and in reflection during steps 34-36.

Table 3. 'Crests' in Clinical Reasoning Task Categories, (all +CF think-alouds, n=30 vs. intermediate think-alouds (+CF and –CF), n=42)

	Contextual Factor				Intermediate Physicians			
step	F	D	M	R	F	D	M	R
1	87%	10%	0%	3%	83%	14%	0%	2%
2	43%	50%	0%	7%	62%	36%	0%	2%
3	60%	37%	3%	0%	48%	48%	2%	2%
4	33%	63%	0%	3%	33%	62%	0%	5%

5	30%	67%	3%	0%	29%	71%	0%	0%
6	40%	57%	0%	3%	48%	52%	0%	0%
7	30%	63%	7%	0%	29%	67%	2%	2%
8	27%	73%	0%	0%	19%	76%	0%	5%
9	20%	80%	0%	0%	19%	79%	0%	2%
10	20%	80%	0%	0%	19%	74%	0%	7%
11	30%	67%	0%	3%	33%	67%	0%	0%
12	20%	73%	7%	0%	21%	76%	2%	0%
13	30%	67%	3%	0%	29%	67%	5%	0%
14	23%	73%	0%	3%	26%	71%	0%	2%
15	33%	63%	3%	0%	38%	55%	7%	0%
16	27%	63%	10%	0%	38%	53%	8%	3%
17	33%	60%	7%	0%	33%	60%	3%	5%
18	25%	68%	7%	0%	31%	67%	3%	0%
19	30%	63%	4%	4%	24%	71%	5%	0%
20	33%	56%	11%	0%	30%	57%	11%	3%
21	30%	67%	4%	0%	32%	65%	3%	0%
22	37%	56%	7%	0%	42%	56%	3%	0%
23	26%	70%	4%	0%	39%	53%	6%	3%
24	37%	52%	11%	0%	40%	51%	9%	0%
25	35%	62%	4%	0%	32%	62%	3%	3%
26	40%	48%	12%	0%	36%	45%	12%	6%

27	36%	50%	14%	0%	39%	50%	7%	4%
28	35%	45%	15%	5%	31%	62%	8%	0%
29	50%	50%	0%	0%	48%	48%	4%	0%
30	26%	58%	16%	0%	25%	50%	21%	4%
31	42%	47%	5%	5%	33%	50%	8%	8%
32	41%	47%	12%	0%	35%	55%	10%	0%
33	24%	59%	18%	0%	25%	65%	10%	0%
34	20%	40%	33%	7%	21%	58%	11%	11%
35	23%	54%	8%	15%	29%	59%	6%	6%
36	15%	38%	46%	0%	18%	76%	0%	6%
37	15%	62%	23%	0%	24%	65%	12%	0%
38	31%	54%	15%	0%	33%	60%	7%	0%
39	38%	54%	8%	0%	40%	60%	0%	0%
40	15%	54%	31%	0%	29%	64%	7%	0%
41	8%	58%	33%	0%	23%	62%	15%	0%
42	17%	67%	17%	0%	15%	62%	15%	8%
43	17%	33%	50%	0%	33%	33%	33%	0%

Discussion

This study investigated physicians' clinical reasoning tasks in an effort to identify if discernable patterns exist. The sequential pattern of clinical reasoning tasks that emerged from this investigation demonstrates that overall, framing tasks are typically the first employed followed by diagnostic, then management. This is consistent with

expectations given the sequential nature of taking a history followed by performing a physical and then generating diagnostic and management decisions. Overall, reflection, though occurring throughout all think-aloud processes, seems to take place closer to the end of the think-aloud process, when present. In fact, based on our analysis, we believe it may occur more frequently once management tasks begin to increase. This runs contrary to our initial hypothesis and may suggest that physicians begin to engage in more thoughtful and effortful processing of reflection prior to establishing a management plan. That is, physicians may reflect on an established diagnostic label and then come up with a plan that is specific to a patient's circumstances and preferences. Empirical evidence supports that this type of iterative reflection contributes to adopting a new perspective that may yield optimal clinical decisions (11). This back and forth movement may also be indicative of physicians moving from higher to lower level cognitive reasoning activities and/or may reflect that management reasoning is less a matter of pattern recognition than arriving at a diagnostic label. The steps that physicians take and their pathway to diagnosis have been identified as important elements in achieving an accurate and timely diagnosis (6). Also, as expected, physicians displayed a variety of paths (different steps) before arriving at diagnostic and management decisions in these cases.

While this analysis captures the general trends and patterns observed, it is important to highlight that this is not a linear process because all tasks, in some magnitude, are in use throughout the think-alouds. This is consistent with situated cognition theory which maintains that clinical reasoning emerges from interactions among the physician, the patient and the environment. For instance, a handful of physicians employed reflection

as their first task, others diagnosis, others framing and some management. This variability suggests the critical role context likely plays in driving clinical reasoning processes, such that each physician proceeds uniquely even when presented with an identical scenario.

Guided by situativity theory, we delved into our emerging pattern to examine the role of contextual factors on clinical reasoning task sequences. Our findings potentially demonstrate the constraints contextual factors pose on cognitive processes, such that the fluidity and flexibility of clinical reasoning, mimicking that of experienced physicians, is only evident in the absence of contextual factors. Perhaps this is an indication that the presence of a contextual factor thwarts both reflective and management activities. Our data suggests that this fluidity may allow for more cognitive flexibility, which leads to a higher rate of framing tasks, a delayed transition into management tasks, and more opportunities to reflect on their patient encounter. Whether this process leads to a more accurate diagnosis deserves further examination. These patterns are in line with previous research findings that support the role of cognitive overload, as a result of contextual factors, and its potential to inhibit cognitive flexibility (4, 7).

Finally, we identified additional evidence of contextual factors indeed disparately impacting clinical reasoning performance. As shown in Figures 2-4 and Tables 2 and 3, the presence of contextual factors has the potential to interrupt the range of clinical reasoning processes. In particular, in the presence of contextual factors, experienced physician patterns are similar to the performance of less experienced physicians. A

plethora of studies have examined the cognitive and metacognitive differences between experts and novices (12-15). Reingold and Sheridan (16) and Taylor (17) are among the many who have studied the differences in perceptions between novices and experts in the medical field. Similarly, Gobet (18) asserts that while medical students begin with a list of possible diagnoses, expert physicians analyze the symptoms to form a diagnosis. Future research could disaggregate task categories into their 26 distinct tasks to further investigate whether experts are focusing on more complex management tasks.

Gobet (18) and Greeno and Simon (19) also argued that experts often rely on general heuristics (pattern recognition) and have the ability to successfully integrate their domain-specific knowledge to other areas. This may help explain the increase in amplitude, or likely cognitive flexibility, that emerged when examining sequence in experienced physician think alouds. The ability to apply patterns and loop in domain-specific knowledge may provide some insight as to why experts seem to "jump" back and forth between clinical reasoning tasks. Additionally, and consistent with the literature, when presented with a potentially more cognitively demanding task (presence of contextual factors), experts have a tendency to employ more robust problem representation where novices rely upon superficial features (20, 21). This is corroborated by the comparative patterns observed herein. Based on our analysis, less experienced physicians seem to perform in manners that mimic a proclivity for "superficial representations," or contextual factors. Conversely, experienced physicians seem to perform in ways that demonstrate that they have more robust representations of a clinical encounter.

Our study was limited by several factors. First, our sequential data was derived from a larger study on clinical reasoning tasks. As such, we employed secondary data that was not designed or collected with the sole purpose of examining patterns and sequences in clinical reasoning tasks. Second, the observed impact of a contextual factor on clinical reasoning task sequencing may be a result of the specific type of contextual factor chosen for the study. Third, the sample breakdown of physician experience was skewed towards the less experienced group (42 vs 18 experts). A more balanced range of experience levels in participants may have yielded more discernible patterns in our comparative analysis. And fourth, the use of think alouds presents unique limitations that prevent a more comprehensive and in-depth examination of participants' thought processes. Participants may not have reflective awareness, or insight, of their own clinical reasoning processes and thus, may not intentionally verbalize these tasks when engaged in a think-aloud process.

These emerging sequences and patterns can be used to inform teaching and assessment of clinical reasoning. Though Juma and Goldszmidt disaggregated the clinical reasoning process into 26 separate teachable tasks, our findings provide further evidence that these tasks mirror the sequence of a traditional patient encounter. As such, instructors can target tasks within any of the four categories (framing, diagnosis, management, and reflection) based on a clinical encounter. For example, if a trainee is having challenges revolving around the diagnosis of a patient, an instructor may want to consider focusing on any of the first three clinical reasoning tasks to help the trainee improve how he/she may be framing the encounter. Extending this example further, if a trainee is having issues related to the patient's management plan, an instructor may choose to initially

encourage the trainee to practice tasks 1-11 in an effort help them build a stronger foundation of evidence prior to establishing a management plan. While additional empirical investigation is needed, providing trainees with such a framework may provide a foundation for establishing more optimal clinical reasoning practices early on in medical education. Previous research demonstrates cognitive flexibility and higher level reflection skills are trainable and can be developed (6).

If in fact contextual factors regress clinical reasoning performance as our findings suggest, this may help instructors design educational experiences, such as simulations, that can be used in continuing medical education. Relying on the concept of deliberate practice (22), this consistent and continued practice in the presence of contextual factors may lead to improved diagnostic accuracy and error reduction.

Our findings provide further insight into the relationship between clinical reasoning and context specificity and how our observed patterns may serve as a useful tool for instruction and assessment of clinical reasoning. This research adds to the existing body of evidence for situated cognition as an appropriate theory for exploring clinical reasoning and informs how future studies might further inform diagnostic and management accuracy. Furthermore, establishing an awareness of the extant sequential patterns promotes considerable progress toward a more complete understanding of clinical reasoning steps. Future research may fruitfully be directed at exploring these task-based patterns and processes to reduce error and enhance diagnostic precision.

The views expressed in this article are those of the authors and do not necessarily reflect the official position or policy of the U.S. Government, Department of Defense, Department of the Navy, or the Uniformed Services University.

This study was supported by a grant from the JPC-1, CDMRP—Congressionally Directed Medical Research Program (No. NH83382416).

References

1. Juma S, Goldszmidt M. What physicians reason about during admission case review. Advances in Health Sciences Education. 2017;22(3):691-711.
2. Schmidt HG, Norman GR, HP B. A cognitive perspective on medical expertise: theory and implications. Academic Medicine. 1990;65:611-21.
3. Eva K. What every teacher needs to know about clinical reasoning. Medical Education. 2005;39(1):98-106.
4. Durning S, Artino A, Pangaro L, van der Vleuten C, Schuwirth L. Context and clinical reasoning: understanding the perspective of the expert's voice. Medical Education. 2011;45(9):927-38.
5. McBee E, Ratcliffe T, Goldszmidt M, Schuwirth L, Picho K, Artino A, et al. Clinical reasoning tasks and resident physicians: what do they reason about? Academic Medicine. 2016;91(7):1022-8.
6. Durning S, Artino A. Situativity theory: a perspective on how participants and the environment can interact: AMEE Guide no. 52. . Medical Teacher. 2011;33(3):188-99.
7. Durning S, Artino A, Boulet J, Dorrance K, van der Vleuten C, Schuwirth L. The impact of selected contextual factors on experts' clinical reasoning performance (does context impact clinical reasoning performance in experts?). Advances in Health Sciences Education. 2012;17(1):65-79.
8. ten Cate O, Durning S. Understanding clinical reasoning from multiple perspectives: A conceptual and theoretical overview. Principles and Practice of Case-based Clinical Reasoning Education: Springer, Cham; 2018. p. pp.35-46.
9. Konopasky A, Ramani D, Ohmer M, Battista A, Artino A, McBee E, et al. It Totally Possibly Could Be: How a Group of Military Physicians Reflect on Their Clinical Reasoning in the Presence of Contextual Factors. Military Medicine. 2019;(in press).
10. Durning SJ, Artino A, Boulet J, La Rochelle J, Van der Vleuten C, Arze B, et al. The feasibility, reliability, and validity of a post-encounter form for evaluating clinical reasoning. Medical Teacher. 2012;34(1):30-7.
11. Zelazo P. Executive function: reflection, iterative reprocessing, complexity, and the developing brain. Developmental Review. 2015;38:55-68.
12. Ball L, Ormerod T, Morley N. Spontaneous analogising in engineering design: a comparative analysis of experts and novices. Design Studies. 2004;25(5):495-508.
13. Bedard J, Chi M. Expertise. Current directions in psychological science. 1992;14(1):135-9.
14. Cross N. Creative thinking by expert designers. The Journal of Design Research. 2004;4(2):123-43.
15. Dixon R. Selected core thinking skills and cognitive strategy of an expert and novice engineer. Journal of STEM Teacher Education. 2011;48(1):7.
16. Reingold E, Sheridan H. Eye movements and visual expertise in chess and medicine. Oxford handbook on eye movements2011.
17. Taylor P. A review of research into the development of radiologic expertise: Implications for computer-based training. Academic Radiology. 2007;14(10):1252-63.
18. Gobet F, Johnston SJ, Ferrufino G, Johnston M, Jones MB, Molyneux A, et al. "No level up!": no effects of video game specialization and expertise on cognitive performance. Frontiers in Psychology. 2014;5:1337.

19. Greeno J, Simon H. Problem solving and reasoning. 2nd ed. Atkinson R, Hernstein R, Lindzey G, Luce R, editors. New York: Wiley; 1988.
20. Hmelo-Silver C, Nagarajan A, Day R. “It's harder than we thought it would be”: A comparative case study of expert−novice experimentation strategies. Science Education. 2002;86(2):219-43.
21. Larkin J, McDermott J, Simon D, Simon H. Expert and novice performance in solving physics problems. Science. 1980;208(4450):1335-42.
22. Ericsson K. Deliberate practice and acquisition of expert performance: a general overview. Academic Emergency Medicine. 2008;15(11):988-94.

Clinical Reasoning in the Medical/Surgical Ward Setting: A Rapid Response Scenario for Residents and Attendings

Megan Ohmer, BS
Graduate Programs in Health Professions Education
Uniformed Services University

Steven J Durning, MD, PhD
Graduate Programs in Health Professions Education
Uniformed Services University

Walter Kucera, MD
Department of Surgery
Walter Reed National Military Medical Center

Matthew Nealeigh, DO
Department of Surgery
Walter Reed National Military Medical Center

Sarah Ordway, MD
Department of Internal Medicine, Division of Gastroenterology
Walter Reed National Military Medical Center

Thomas Mellor, DO
Department of Internal Medicine, Division of Gastroenterology
Naval Medical Center San Diego

Jeffrey Mikita, MD
Department of Simulation
Walter Reed National Military Medical Center

Anna Howle, MAC
Department of Simulation
Walter Reed National Military Medical Center

Sarah Krajnik, RN, BSN
Department of Simulation
Walter Reed National Military Medical Center

Abigail Konopasky, PhD
Graduate Programs in Health Professions Education
Uniformed Services University

Divya Ramani, MS
Graduate Programs in Health Professions Education
Uniformed Services University

E. Matthew Ritter, MD, MEd
Department of Surgery
Walter Reed National Military Medical Center

Alexis Battista, PhD
Graduate Programs in Health Professions Education
Uniformed Services University

Corresponding Author:
Megan Ohmer, BS
Research Assistant
Graduate Programs in Health Professions Education
Department of Medicine
F. Edward Hébert School of Medicine
Uniformed Services University of the Health Sciences
4301 Jones Bridge Road
Bethesda, MD 20814-4712
megan.ohmer@usuhs.edu
(301) 295-9540

Structured Abstract: 250
Word Count: Words (less references): 4286
Tables: 4
Figures: 0
References: 36
Associated Appendices: 14

Disclaimer:
The opinions and assertions expressed herein are those of the authors and do not necessarily reflect the official policy or position of the Uniformed Services University or the Department of Defense.

Conflict of interests:
The authors have no conflicts of interest to declare or any financial interest in any commercial product, service, or organization providing financial support for this research.

Grant Support:
Developing Assessment Tools to Better Understand the Mechanisms of Clinical Reasoning in Military Medical Simulation
JPC - 1, CDMRP - Congressionally Directed Medical Research Programs
Award #: NH83382416
Amount: $1,535,000.00

Ethical Approval:
This research protocol was reviewed and approved by the Uniformed Services University of the Health Sciences Office of Human Subject Protection institutional review board (IRB). In the conduct of research where humans are the subjects, the investigators adhered to the policies regarding the protection of human subjects as prescribed by Code of Federal Regulations (CFR) Title 45, Volume 1, Part 46; Title 32, Chapter 1, Part 219; and Title 21, Chapter 1, Part 50 (Protection of Human Subjects).

Acknowledgments:
We gratefully acknowledge the support and guidance from Drs. Paul Hemmer and Jeff LaRochelle while refining these scenarios prior to their implementation. We also want to thank our standardized participant, Michael Betteridge, and the Walter Reed Simulation Center operations team for their advice and feedback throughout development and implementation.

Abstract (250/250)

Introduction: This resource describes the development, implementation and evaluation of a scenario-based simulation to assess physicians' clinical reasoning processes when presented with a patient in the medical surgical ward setting requiring resuscitative care resulting from a tension pneumothorax.

Method: The authors drew upon participatory design while integrating select simulation design strategies, including diagnostic ambiguity, contextual factors, and rising patient acuity to increase complexity. The scenario was designed to be used with an open-ended written exercise and a think-aloud reflection protocol, also included, to elicit clinical and management reasoning and reflection on that reasoning. Descriptive statistics were used to analyze results from 20 intern through attending physicians with specializations in family or internal medicine or surgery.

Results: Half of physician participants identified the correct leading diagnosis (*n* = 10), tension pneumothorax, while the second most common leading diagnosis listed was a pneumothorax (*n* = 8). Physician participants considered a total of 19 additional differential diagnoses. Regarding treatment, participants requested additional help and provided supportive care (*n = 4*), performed a needle thoracostomy (*n* = 10) or a tube thoracostomy (*n* = 3), or both (*n* = 3). Self-reported cognitive load ranged from four to ten (*m* = 7.15) (on a scale from one to ten). Scenario authenticity was rated between three and five (*m* = 4.13) (out of five).

Discussion: This resource includes an inpatient scenario-based simulation and supporting materials to prompt diagnostic and management reasoning and reflection upon that reasoning. It can be used to support interns through attending physicians.

Educational Objectives:

As a result of participating in this scenario, participants will:

1. Practice engagement in clinical behaviors in the inpatient setting, including gathering and analyzing information and evidence and interpretation of that evidence to formulate a differential diagnosis, a problem list and a leading diagnosis with supporting evidence.
2. Practice planning and performing a focused physical exam to further support clinical reasoning.
3. Provide immediate resuscitative care to a patient suffering from a tension pneumothorax and then practice developing and communicating a future management plan.
4. Reflect on their own diagnostic and management reasoning efforts and strategies by thinking aloud and observing their own video-recorded simulation.

Introduction and Background

Clinical reasoning is a central activity in making *accurate diagnoses* and includes the activities of gathering and synthesizing information, interpreting data (e.g., responses to diagnostic questions, interpreting laboratory or imaging studies), generating multiple differentials, refining those initial hypotheses, and developing additional testing or future therapeutic plans. Errors in clinical reasoning processes are viewed as a primary contributing factor to diagnostic error, which includes instances where the correct diagnosis is missed, delayed, or wrong (Graber, 2012). According to Graber(2012), among the primary specialities, such as family and internal medicine, emergency medicine and general surgery, diagnostic errors may occur in up to 10-15% of patient visits. Research by Newman-Toker and others suggests that diagnostic error may result in up to 40,000 patient deaths and 150,000 harms to patients annually (Balough, Miller & Ball, 2015; Newman-Toker and Macary, 2013; Newman-Toker and Provnost, 2009).

A retrospective analysis conducted by Bishop and colleagues indicates that among inpatient errors, surgical errors were more prevalent; however, the second most common source of errors in the inpatient setting included diagnostic errors at a rate of 21%. (Bishop, Ryan & Casalino, 2011). Factors that are believed to influence a physician's clinical reasoning processes include, deficits in cognitive processes (e.g., System I, System II errors) (Graber, Franklin & Gordon, 2005), deficits in medical knowledge (Norman, 2005; Norman, Monteiro, Sherbino, Ilgen et al., 2017), inappropriate selectivity (i.e., not being selective in the information one gathers or follows up on)(Zwan et al., 2013), affective variables (e.g., how the physician feels about the patient, or themselves) (Croskerry, Abbass, and Wu, 2010), and contextual factors (e.g., physician factors, patient factors, environmental factors)(Durning et al., 2012; Eva, 2005e). Recent research indicates that physicians along a continuum of practice (i.e., interns through attendings) could benefit from continued focused practice in clinical reasoning processes [Konopasky et al., unpublished manuscript). Thus, there is a need for educational and assessment resources that can support the practice and assessment of the complex processes of clinical reasoning and that can be used with physicians along the fuller continuum of practice in the inpatient setting.

Audience and Contribution

Building on prior instructional design and research efforts examining clinical reasoning in the outpatient setting [Battista et al., in press], we aimed to create a suite of resources to examine clinical reasoning processes in the inpatient medical-surgical ward setting depicting a patient with a time sensitive, life threatening condition, namely a developing tension pneumothorax following blunt trauma to the chest. This resource was originally designed to formatively assess and research physicians' (interns through attendings) clinical reasoning processes in the inpatient setting; however, the resource can also be used to support the formative development of, assessment and remediation of, and deliberate practice of clinical reasoning. This article describes the scenario's development, testing and initial implementation evaluation. This article does not report the implementation of the think-aloud reflections which are reported in detail in Battista et al. (in press).

This resource adds to the diverse and growing body of resources in MedEdPORTAL addressing clinical reasoning. For example, MedEdPORTAL provides numerous resources focusing on improving or remediating clinical reasoning, including debiasing strategies (e.g., Weinstein & Pinto-Powell, 2016; Weinstein et al., 2017; Levin, Cennimo, Chen & Lambda, 2016; Pitre, 2016; Daniel et al., 2017) while others include cases or scenarios that individuals can engage in for practice (see for example, Wittler, Obrien & Masneri, 2018; Maniaci, Simon & La Rosa, 2015; Lupi, Oliva & Splinter, 2014; Barrat & Obeso, 2012). This resource adds to these by: a) providing a scenario that addresses clinical reasoning in the inpatient medical-surgical ward context (as opposed to the emergency department or outpatient setting), b) coupling the scenario with a reflection protocol to elucidate clinical reasoning processes, c) being adaptable to both medicine and surgical specialities, and d) being applicable to physicians ranging from intern through attending (versus medical students).

Method

Resource Development

Resource goals: Drawing on previous efforts to examine and assess clinical reasoning processes (Battista et al., In Press) we developed a scenario-based simulation to elicit clinical reasoning in a straightforward inpatient experience. We coupled the scenario with written and verbal reflection tools to evaluate physicians' reasoning. Given that our prior resources emphasized patients with low to moderate levels of acuity we sought to develop a resource to examine clinical reasoning when the patient's acuity was higher.

Scenario-based simulation rationale: Our strategy in employing a live scenario-based simulation to elicit clinical reasoning was multifold. Simulations are commonly used to present learners or study participants with a high risk yet low frequency clinical situation because simulation offers the advantages of predictability and no risk to actual patients (McGaghie et al., 2005; Gaba 2007). Simulations are also commonly used to study diverse individual- or team-level behaviors (Tschan), communication (Konopasek, Kelly & Bylun, 2014) and decision making (Hallin et al., 2016) or clinical reasoning (Battista et al., In press). Scenario-based simulations, which employ a narrative and often require participants to identify and resolve a problem, are intended to closely approximate the actual clinical setting they represent both by providing interactions with common tools and artifacts (e.g., stethoscope, exam table) and common clinical roles (e.g., patient, other healthcare professionals) (Gaba, 1996; Dieckmann et al. 2007; Battista, 2017).

Design procedures for the scenario: We used participatory instructional design processes, which encourage the inclusion and integration of the perspectives of diverse stakeholders. (Konings et al., 2010) The process included three stages: initial design (e.g., determining scenario goals, identifying stakeholders), preliminary testing (e.g., read throughs, rehearsals) and implementation with evaluation (e.g., analysis of implementation processes and participants' performance). We refer the reader to [Battista et al., In Press; Konings et al., 2010) for a more detailed description of participatory design processes.

Design features used to develop the scenario: We used several strategies to design a scenario that could be undertaken by interns, residents or attendings representing differing specialities, including family medicine and internal medicine and surgery. First, we selected an admission diagnosis (cellulitis with an abscess drained in the emergency department) that could plausibly be admitted to either a medical or surgical team. To accommodate differences in participants' experiences, we designed the scenario with more than one branching option because we reasoned that some participants might not be comfortable performing a needle decompression independently while others might proceed to a needle decompression (see Appendix J for branching diagram). We then used three strategies to increase scenario complexity, namely diagnostic ambiguity, contextual factors, and increased patient acuity, to provide a sufficiently complex case for interns through attending physicians.

To design scenarios with *diagnostic ambiguity*, defined as a series of symptoms and findings which could suggest more than one diagnosis, we drew from Tschan and colleagues' work [Tschan et al., 2009], developing a history of present illness -- pain and trauma associated with a fall -- that could have been caused by more than one source, such as simply tripping and falling or syncope, for example (see Appendix N for a detailed cue - hypothesis matrix adapted from Elstein)(Elstein, 1978).

We also included a *contextual factor*--reporting to the bedside of a patient unknown to the physician participant--to further introduce uncertainty. According to recent literature, contextual factors may influence clinical reasoning performance, potentially introducing significant unwanted variance (error) in patient care [Durning et al., 2011]. When contextual factors are introduced, context specificity may occur—the phenomenon whereby a physician sees two patients with identical symptoms and findings and yet comes to two *different* diagnostic decisions[Durning et al., 2011]. This emerging literature on contextual factors suggests that both residents and attending physicians are affected by them and thus the uncertainty they introduce would be expected to impact all participants in the scenario (Durning et al., 2011; Durning et al., 2012).

Preliminary findings from our prior clinical reasoning research in the outpatient setting indicated that increased patient *acuity* resulted in participants reporting the scenario as being more complex [Battista et al., In Press], so we designed the scenario so that the patient's condition would deteriorate quickly as their tension pneumothorax worsened.

Rationale and Description of Selected Measures

The scenario development process was accompanied by the selection of measures and reflection tools to help participants make their reasoning more explicit and, thereby, to support analysis of clinical reasoning. Measures include: an open-ended post-encounter form (PEF) to elicit diagnostic and management reasoning and a verbal think-aloud protocol to elicit reflections on reasoning. To determine if the scenario was sufficiently challenging we also included single-item cognitive load and a perceived authenticity question to determine participant satisfaction with the scenario. We describe these measures in greater detail next.

Post encounter form: The PEF measure (Appendix K) asks participants to provide a) any additional questions they would ask, b) any additional exam actions they would take, c) a problem list, d) a differential diagnosis, e) a leading diagnosis, f) supporting evidence for that diagnosis, and f) a treatment plan (Durning et al., 2012). This form was given to participants immediately after they completed the scenario and participants had up to 30 minutes to complete it (see Appendix A for workflow).

Think-aloud protocol: The retrospective "think-aloud protocol " is a strategy that can be used to help elicit insight into an individual's cognition and experience while also strengthening their own learning (see Appendix C for think-aloud warm-up and detailed implementation instructions) (Ericsson; Burbach). This process involved asking participants to watch a video-recording of their own performance and provide a stream of consciousness reflection on what they were thinking while partaking in the scenario. The think aloud was conducted immediately following completion of the PEF (see Appendix A for workflow).

Cognitive load: Following completion of the scenario and the PEF (See Appendix A for workflow), we assessed participants' self-reported cognitive load using a single item adapted from Plass and colleagues asking participants to "please rate your invested mental effort after completing the post-encounter form" on a scale ranging from one (very low mental effort) to ten (very high mental effort) (Appendix L).

Scenario Authenticity: Finally, we developed a single item question that asked participants to rate the authenticity of the scenario on a scale ranging from one (not at all authentic) to five (very authentic) (Appendix M).

Scenario Procedures and Logistics

Scheduling logistics: For each scheduled date we requested two rooms in the simulation center. The first room was used to allow participants to complete the think-aloud warm up prior to participating in the scenario and to complete the PEF and re-watch their own video-recorded performance while thinking aloud afterwards. No special setup was required for the first room. The second room mimicked a typical medical/surgical ward hospital room, including a hospital bed, an IV pole, a chair, a sink, and a headwall with oxygen hookup and call bells. We asked the simulated participant (SP) portraying the patient to arrive approximately 30-45 minutes prior to the session start for moulage application (see Appendix H for moulage images and instructions).

Simulated stethoscope: Participants used an AURiS simulation stethoscope which mimicked abnormal breath sounds in the standardized patient; however, others should be able to utilize alternative simulated stethoscope models depending on what is available in their local simulation center. A complete supply list is included in the appendices (Appendix G). We also created a mock emergency department medical record (Appendix I) that was available for participants to review if they desired.

Hybrid simulation strategy for needle or chest thoracostomy: To support participants' performance of a needle or tube thoracostomy we utilized a hybrid simulation strategy whereby the simulated participant portrayed the patient but a chest model was used to allow participants to perform these invasive procedures. We utilized the TraumaMan® surgical simulator because this was the model available in our designated simulation center; however, others should be able to use any variety of chest models that are available to them.

Staffing requirements: We scheduled three team members, in addition to the SP portraying the patient role, for each session. Their roles were as follows:

- The first team member, a trained research associate, was responsible for greeting each participant, ensuring that they were oriented to the simulation and think-aloud activities and that they completed all of the steps of the session. This team member also acted as the patient-care tech, assisting in the simulation.
- The second team member with medical knowledge and trained to portray the primary nurse, remained in the room for the duration of the simulation and assisted the participant as a typical medical/surgical nurse would, as well as addressing any technical issues that may arise during the scenario.
- The third team member portrayed the first rapid response team (RRT) member to arrive to the patient's room, eliciting information from the participant, ending the scenario, and conducting the retrospective think aloud.

Video recording and video playback during think alouds: To support the replaying of participant videos during their think alouds, we video recorded each scenario using three portable video cameras fitted with removable SD cards. We used multiple cameras so that we had back-up copies in case the primary camera failed. Following each scenario, while the participant completed the PEF, a team member removed the SD card and connected it to a computer with an SD card reader so that the participant could then re-watch their video.

Think alouds: The think-aloud protocol was conducted by a trained research associate who remained with the participant while they viewed their performance and thought aloud. Think alouds were audio recorded using two digital voice recorders--one recorder acting as a backup (see Appendix C for detailed instructions on how to conduct a think aloud).

Participant procedures: On the scheduled day, participants were oriented to the session workflow (see Appendix A), simulation expectations (Appendix B) and the think aloud (Appendix C). Because we recruited participants of varying experience and specialty, we designed the scenario with multiple branch points and instructed participants to practice within the limitations of their comfort level (Appendix B).

To start the scenario, participants:

1. received a phone call (using one of the researcher's cell phones) from the "primary nurse" calling to report the patient's fall (see Appendix F, page 1, for primary nurse scripted opening); and
2. were directed to the patient's room where they had up to 20 minutes to complete their assessment and any interventions that might be required (there was no penalty for finishing early or being stopped before completion); and
3. the scenario was stopped once the participant gave a brief report to the first member of the rapid response team. The rapid response team member exercised their best judgement in determining when they had received enough information.

Observation for and management of strugglers: Because we included participants with a wide range of experience we anticipated that some participants might struggle or delay escalating care and seeking additional help. Participants were observed by the primary nurse and one of the team members not present in the room (the rapid response team member) for extensive delays. Delays often manifested as instances where participants remained focused on their assessment while failing to call for help or express concern or consideration of the patient's

deterioration. In these cases, the primary nurse was instructed to verbally indicate that they were concerned about the patient's condition and were going to call a rapid response.

Following the scenario performance, participants:

4. completed the PEF (Appendix K);
5. completed the cognitive load question (Appendix L);
6. re-read the instructions for thinking aloud (Appendix C);
7. watched their own video-recorded performance while thinking aloud (Appendix C); and
8. completed the scenario-authenticity question (Appendix M).

Participants' total time to complete the scenario, the related post encounter form and think aloud, and the other informational questionnaires ranged between one hour and 15 minutes to one hour and 30 minutes (completion of the PEF accounted for the majority of the variability in total time as completion ranged from 15 - 30 minutes).

Optional feedback: While these scenarios and reflection protocols were initially intended to support our research, we recognized that participants could benefit from feedback from a dedicated study team physician with expertise in treating patients with similar conditions. Therefore, following each simulation we notified the participants verbally and by follow-up email that individual feedback on their performance was available if they so desired.

Casting, Training and Quality Improvement

SP casting and training: We sought an SP similar to our designed role in age and body habitus (a middle-aged male of average weight and stature). The SP trainer used a series of possible questions in a rehearsal guide (Appendix E) developed from our initial design sessions; however, SPs were not required to adhere verbatim to the guide. Rather, our goal, given that we initially used these resources for research and conducted scenarios over several months, was to support implementation fidelity. SPs were instructed to provide information if prompted and minimize volunteering.

Data Analysis of Preliminary Findings

To determine if our developed scenario enabled participants to consider more than a single diagnosis, i.e., evidence of scenario ambiguity [Tschan, Semmer, Gurtner et al., 2009], and if the scenario reliably led participants to perform the anticipated treatment actions, we conducted a content analysis and calculated descriptive statistics for frequency counts of the post encounter forms (PEFs) (i.e., free text questions about leading and differential diagnoses, problem list, supporting evidence and management decisions (Appendix I))[12]. To determine if the scenario was appropriately complex for diverse participants (interns through attending), we calculated descriptive statistics of participants' self-reported cognitive load (Appendix J) and their perceptions of scenario authenticity (Appendix K).

Results

Participants in this sample were 20 internal medicine, family medicine and general surgery physicians, six female and 14 male. Twelve were resident physicians (six from post graduate year 1 [PGY 1], six from PGY 2-3), one participant was a fellow (hereafter grouped with attendings), and seven were attendings (those having completed their initial residency).

Table 1. Participant Demographics

		Intern	Resident (PGY 2-4)	Attending
Gender				
	Male	5	3	6
	Female	1	3	2
Specialty				
	Internal Medicine	5	6	5
	Family Medicine	0	0	1
	General Surgery	1	0	2

Analysis of the PEF

Leading diagnosis and supporting evidence: Participants listed a total of four different diagnoses as their leading diagnosis as measured by the PEF (Table 2).

Table 2. Most Common Leading Diagnoses

Leading Diagnosis	**Frequency of listed leading diagnosis for interns (*n* = 6)**	**Frequency of listed leading diagnosis for residents (*n* = 6)**	**Frequency of listed leading diagnosis for attendings (*n* = 8)**
Tension Pneumothorax	4	1	5
Pneumothorax	1	5	2
Hemothorax	0	0	1
Sepsis/Septic Shock	1	0	0

Note: Bold indicates the most frequently selected leading diagnosis.

The PEF revealed that the most common *supporting evidence* participants listed included hypotension (*n* = 17), recent chest trauma (*n* = 15), absent or decreased breath sounds in the right lung fields (*n* = 14), tachycardia (*n* = 13), hypoxia (*n* = 10). Participants who conducted a needle decompression also noted the improvement in vital signs and mental status (*n* = 6 for each) as supporting evidence. Interns (PGY-1) listed between four and six items of supporting data (*m* = 4.67), residents (PGY-2, -3) listed between four and seven (*m* = 5.0) and attendings listed between one and eight *(m* = 4.75).

Differential diagnoses: Participants considered a total of 19 independent differential diagnoses (Table 3). The most common differentials included tension pneumothorax (*n* = 13), pulmonary embolism (*n* = 11), pneumothorax, hemothorax, rib fracture, and sepsis (*n* = 9 for each). These appeared to differ by PGY: interns considered nine independent differential diagnoses, residents considered ten, and attendings listed 15. The number of differential diagnoses listed by each participant ranged from three to seven (*m* = 4.2). Interns listed between three and five differentials (*m* = 3.5), residents listed between three and six differentials *(m* = 4.16) and attendings listed between three and seven differentials (*m* = 4.63).

Table 3. Most Common Differential Diagnoses Considered

Differential Diagnosis	Frequency of listed differential diagnoses for interns (n = 6)	Frequency of listed differential diagnoses for residents (n = 6)	Frequency of listed differential diagnoses for attendings (n = 8)
Tension Pneumothorax	5	2	6
Pneumothorax	1	5	3
Hemothorax	0	4	5
Rib Fracture/Flail Chest	1	3	5
Sepsis	3	4	2
Cardiac Tamponade	2	1	3
Pulmonary Embolism	4	3	4
Angina/ACS/MI	3	1	3

Note: Frequency count exceeds 20 because participants were not limited in the number of differential diagnoses they could list. Additional diagnoses listed that received two or fewer mentions included: Pulmonary contusion, anaphylaxis, HHS/DKA, underlying lung disease (leading to pneumothorax), malignancy (leading to pneumothorax), stroke, obstructive shock, pneumonia, liver injury, diaphragm injury, and cardiogenic syncope

Management considerations. As participants had the opportunity both to administer interventions during the scenario and list other management considerations on the PEF, we examined the PEFs for management reasoning. Per ATLS guidelines as well as input from our subject matter experts, participants who either performed a needle thoracostomy followed by a tube thoracostomy in the scenario OR indicated that those interventions are necessary on their PEF were considered correct as participants were instructed to only practice within their comfort level. Ten participants in our study were graded "correct" according to these standards. Table four provides a detailed description of participants' actions. The participants who did not perform a needle or tube thoracostomy did call for help or activated the rapid response team and provide supportive care in the interim, to include supplemental oxygen and IV fluids.

Table 4. Participant Activities and PEF Responses Related to Treatment

	Called for help/rapid response only	Both needle and tube thoracostomies	Needle thoracostomy only	Tube thoracostomy only
Interns				
-During Scenario	2	1	2	1
-Noted on PEF	X	X	X	X
Residents				
-During Scenario	1	0	4	1
-Noted on PEF	X	X	X	X
Attendings				
-During Scenario	1	2	4	1
-Noted on PEF	X	X	X	X

The most frequent study requested in the PEF was a chest x-ray (*n* = 16); three participants requested an electrocardiograph and four stated they would consider chest computed tomography when the patient was more stable. Participants also mentioned interventions that they would perform in the PEF. The most common was placing a chest tube (*n* = 13), followed by performing a needle decompression (*n* = 9). Participants also stated that they would administer supplemental oxygen (*n* = 8), continue antibiotic treatment, transfer the patient to the intensive care unit (*n* = 7 for each), administer pain medication (*n* = 6) or intravenous fluids (*n* = 5), and perform more invasive airway management (*n* = 4). Other management tasks mentioned were placing the patient on continuous cardiac monitoring, pulmonary toilet and incentive spirometry use, thoracic surgery consult and obtaining labs (complete blood count, coagulation studies, troponins, lactate, arterial blood gas, and blood cultures).

Cognitive Load

Participants' self-reported cognitive load for this scenario ranged from four to ten on a scale of one to ten (*m* = 7.15). While the sample was too small for significance testing, we noted that PGY-1 interns found this scenario to require more mental effort (*m* = 7.83) than residents (*m* = 7.17) or attendings (*m* = 6.63).

Participant Ratings of Scenario Authenticity

Participants generally rated the case as being authentic, with a mean of 4.13 (on a scale of one to five). While there was not enough power to test statistically, we noted that attendings rated the authenticity slightly lower (*m* = 3.94) than interns (*m* = 4.0) or residents (*m* = 4.5).

Discussion

The findings from the implementation evaluation indicate that most participants (*n* = 10) correctly identified the leading diagnosis of tension pneumothorax while, of the remainder, eight participants indicated that the patient suffered from a pneumothorax, one diagnosed hemothorax, and one diagnosed septic shock. Based on participant actions during the scenario, we hypothesize that many of the participants who listed "pneumothorax" (without the specifier of "tension") may have done so because at the time participants completed the form the tension aspect of the pneumothorax was currently treated by chest tube or needle decompression. Participants primarily relied upon the supporting evidence of hypotension, recent trauma, tachycardia and dyspnea as evidence of the patient's condition, suggesting good face validity. Of the 19 participants, most treated the tension pneumothorax by performing a needle decompression (*n* = 10) or a combination of a needle decompression and placing a chest tube (*n* = 3). The remaining participants identified the need to escalate care and asked for additional help, most often by requesting assistance from the rapid response team, which further suggests good face and construct validity. Lastly, although most participants identified the correct leading diagnosis, they also considered numerous (*n* = 19) differential diagnoses, thus suggesting that our strategy to introduce complexity using diagnostic ambiguity may have contributed to this outcome.

In regards to examining if the scenario provided sufficient challenge (measured by cognitive load), the findings suggest that all participants found it at least moderately challenging and that interns and residents reported a greater cognitive load compared to attendings. A similar pattern was also noted in the authenticity ratings, where all participants found it at least moderately authentic and interns and residents (PGY 2-4) rated the authenticity higher than attendings.

Reflections on Development and Implementation

Here we share a few reflections on the design and implementation analysis. First, participatory design methods help support efforts to anticipate and prepare for the actions of participants, which in turn, helps minimize instances where the simulation team is uncertain as to how to respond to a participant's actions. This consistency in implementation is especially important as our use of these scenarios was intended for research and thus helps minimize the possibility that our findings and later inferences are influenced by too much variation in implementation.

We also noted that the use of the simulated stethoscope caused some confusion for some participants despite having an orientation opportunity prior to scenario participation. To address confusion during the scenario and to maintain scenario flow, the primary nurse provided additional guidance during the scenario when needed. Notably, this may have contributed to some participants' lower ratings of scenario authenticity. To further improve the integration of a simulated stethoscope in the future we recommend that participants be provided with an orientation that includes verbal instructions and an opportunity to practice its use if time and staffing permit.

Lastly, our strategy to induce complexity, namely, responding to the bedside of an unknown patient, resulted in three participants expressing that they felt this strategy was unrealistic even though we set the start of this scenario to depict the patient has having just arrived on the floor. This, too, could have also contributed to some participants' lower ratings of authenticity. Depending on the context of other organizations employing this scenario, we recommend that users consider building in a mock sign-out to mitigate this issue.

Limitations

Limitations of the analysis include our use of a small sample to make generalizations beyond this group. Limitations to the scenario's implementation and perceived authenticity may include the potential challenges participants reported in working with a simulated stethoscope and in responding to the bedside of an unknown patient.

Conclusion

This resource includes an inpatient scenario-based simulation and supporting materials to prompt and analyze diagnostic and management reasoning and reflection upon that reasoning, The scenario can be utilized independently or coupled with the additional reflection tools associated with this resource. Furthermore, the resource can be used to support interns through attending physicians who represent specializations in internal medicine, family medicine or general surgery.

References

Appendices

Scenario Implementation

A. Clinical reasoning in the medical/surgical ward setting participant workflow diagram
B. Participant expectations and instructions
C. Think-aloud instructions and warm up
D. Standardized patient case
E. Standardized patient rehearsal guide
F. Scenario storyboard
G. Supply list
H. Moulage images and instructions
I. Mock emergency department medical record
J. Tension pneumothorax branching diagram

Measures

K. Post encounter form (PEF)
L. Cognitive load question
M. Scenario authenticity question
N. Cue-Hypothesis Matrix

Clinical Reasoning in the Primary Care Setting: Two Scenario-Based Simulations for Residents and Attendings

Alexis Battista, PhD
Graduate Programs in Health Professions Education
Uniformed Services University

Abigail Konopasky, PhD
Graduate Programs in Health Professions Education
Uniformed Services University

Divya Ramani, MS
Graduate Programs in Health Professions Education
Uniformed Services University

Megan Ohmer, BS
Graduate Programs in Health Professions Education
Uniformed Services University

Jeffrey Mikita, MD
Department of Simulation
Walter Reed National Military Medical Center

Anna Howle, MAC
Department of Simulation,
Walter Reed National Military Medical Center

Sarah Krajnik, RN
Department of Simulation
Walter Reed National Military Medical Center

Dario Torre, MD
Graduate Programs in Health Professions Education
Uniformed Services University

Steven J Durning, MD, PhD
Graduate Programs in Health Professions Education
Uniformed Services University

Corresponding Author:
Alexis Battista, Ph.D.
Assistant Professor of Medicine
Graduate Programs in Health Professions Education
F. Edward Hébert School of Medicine
Uniformed Services University of the Health Sciences
4301 Jones Bridge Road
Bethesda, MD 20814-4712

alexis.battista.ctr@usuhs.edu
(301) 295-9540

Disclaimer:
The opinions and assertions expressed herein are those of the author(s) and do not necessarily reflect the official policy or position of the Uniformed Services University or the Department of Defense.

Conflict of interests:
The authors have no conflicts of interest to declare or any financial interest in any commercial product, service, or organization providing financial support for this research.

Grant Support:
Developing Assessment Tools to Better Understand the Mechanisms of Clinical Reasoning in Military Medical Simulation
JPC - 1, CDMRP - Congressionally Directed Medical Research Programs
Award #: NH83382416
Amount: $1,535,000.00

Ethical Approval:
This research protocol was reviewed and approved by the Uniformed Services University of the Health Sciences Office of Human Subject Protection institutional review board (IRB). In the conduct of research where humans are the subjects, the investigator(s) adhered to the policies regarding the protection of human subjects as prescribed by Code of Federal Regulations (CFR) Title 45, Volume 1, Part 46; Title 32, Chapter 1, Part 219; and Title 21, Chapter 1, Part 50 (Protection of Human Subjects).

Acknowledgments:
We gratefully acknowledge the support and guidance from Drs. Lou Pangaro, Paul Hemmer, Elexis McBee, Jeff LaRochelle and Luke Surry while refining these scenarios prior to their implementation. We also want to thank our standardized participants, Claude Stark, Chris Morrow and Lindsay Williams and the Walter Reed Simulation Center for their advice and feedback throughout development and implementation. We also thank the reviewers who provided thoughtful and invaluable feedback on this manuscript and accompanying appendices.

Structured Abstract

Introduction: This resource describes the development and implementation of resources medical educators or researchers can use for developing or analyzing resident through attending physicians' clinical reasoning in an outpatient clinic setting. The resource includes: a) two scenario-based simulations (i.e., diabetes, angina), implementation support materials, an open-ended post-encounter form, and a think-aloud reflection protocol.

Method: We designed two scenarios with potential case ambiguity and contextual factors to add complexity for studying clinical reasoning. They are designed to be used prior to an open-ended written exercise and a think-aloud reflection to elicit reasoning and reflection. We report on their use in a research context but developed them to be used in both educational and research settings.

Results: Twelve physicians (5 interns, 3 residents, and 4 attendings) considered between three and six differential diagnoses for the diabetes scenario (*m* = 4.0) and between three and nine (*m* = 4.3) differentials for angina. In think-aloud reflections, participants reconsidered their thinking between zero and 14 times (*m* = 3.5) for diabetes and zero and 11 (*m* = 3.3) times for angina. Cognitive load scores ranged from four to eight (out of ten; *m* = 6.2) for diabetes and five to eight (*m* = 6.6) for angina. Participants rated scenario authenticity between four and five (out of five).

Discussion: The potential case content ambiguity along with the contextual factors (e.g., patient suggesting alternative diagnoses) provide a complex environment in which to explore or teach clinical reasoning.

Educational Objectives:

As a result of participating in these scenarios, participants will:

1. Practice gathering and analyzing information and evidence and interpretation of that evidence to formulate differential diagnoses, and forming a problem list and a leading diagnosis with supporting evidence.
2. Practice planning and performing a focused physical exam to further support clinical reasoning.
3. Practice developing and communicating a management plan.
4. Reflect on their own clinical reasoning efforts and strategies by thinking aloud and observing their own video-recorded simulation.

Introduction and Background

Clinical reasoning includes the gathering and synthesizing of information, interpreting data (e.g., patient's responses to diagnostic questions, lab or radiologic findings), generating and refining hypotheses, and problem representation or the use of illness scripts[1-5]. Clinical reasoning is vital to making an accurate diagnosis, eliciting appropriate management and the development of efficient therapeutic plans[1,6,7]. Research examining clinical reasoning suggests it is a complex activity that relies on several factors, including the physician's cognitive processes, knowledge derived from formal and informal experiences, and prior practice experiences (e.g., prior exposure to similar patients)[5,7].

Efforts to assess clinical reasoning use a variety of strategies: among some of the most common are: multiple choice questions[8], case-based learning[9], the integration of think-aloud reflections with video-based scenarios[10] or virtual patient scenarios[11]. In many of these examples, participants are asked to imagine themselves as the hypothetical participant rather than engaging in their own clinical encounter. Live scenario-based simulations are also reported, however, less frequently[12-14]. Moreover, save one example, [Burbach[15]], none of these scenario-based approaches are paired with a free-text, open-ended approach to assessment and reflection, which can offer deeper understanding of the *process* of reasoning[16,17].

In addition, most of these efforts to support clinical reasoning are designed to support individuals still in their undergraduate training rather than health care professionals' learning and development *throughout* their career, something recommended by the recent National Academies of Science report on improving diagnostic efforts[6].

We aimed to create scenarios where we could examine how physicians with a range of experience levels organized their interview, physical exam, diagnostic ideas, and management choices when engaging with a single patient (portrayed by a standardized patient [SP]). We considered that scenario-based simulations, which use a narrative to guide participants' engagement as they address a problem that needs to be explored or resolved[18], would encourage physician performance that would be similar to the actual clinical setting while allowing us to control for the known leading and differential diagnoses. We also considered that these scenarios would provide physician participants with the opportunity to engage in many of the component activities associated with clinical reasoning (e.g., information gathering, interpretation of diagnostic information, hypothesis generation, management plans).

Several authors argue that scenario-based simulations like these are ideal for exploring the complexity of clinical practice, such as clinical reasoning[3,19,20]. For example, Elstein and colleagues describe how they utilized scenario-based simulations to conduct an in-depth descriptive analysis of physicians' behaviors while engaging a SP[3]. Furthermore, Dieckmann and colleagues argue that scenario-based simulations are complex social endeavors that support interactions among health professionals (e.g., medical doctors), simulated participants (e.g., SP) and other culturally relevant devices, such as diagnostic equipment[19]. Kneebone and colleagues suggest simulations present participants with contexts that support the development of skills and knowledge within a context that represents many of the elements of professional clinical practice[20]. And the findings of a more recent descriptive analysis of scenario-based simulations suggests that they provide participants with an opportunity to *make sense* of a clinical situation because they support activities such as information gathering (e.g., diagnostic questioning, interpreting diagnostic findings) as well as carrying out patient management activities[21].

The purpose of this article is to describe the development, testing, design improvements, and implementation of two *live* scenario-based simulations (i.e., new onset diabetes, coronary artery disease presenting with angina), together with an open-ended written exercise and a think-aloud reflection protocol. Here we report on a single study of their use in a research context but we developed them to be used in both educational and research settings. This suite of resources can be used to support researching or teaching resident *and* attending physicians' clinical reasoning in an outpatient clinic setting.

Audience and Contribution

This resource was designed to assess the clinical reasoning of physicians with a range of experience and ability (i.e., residents to attendings). This article describes the strategies we used to develop and test scenarios with the expressed intent of integrating diagnostic ambiguity (where a series of signs and symptoms could be attributed to more than one diagnosis)[14] and contextual factors (referring to factors that may interact, such as patient, physician, and setting factors)[10,22] as ways to increase scenario complexity. Furthermore, pairing the scenarios with two different reflective tools (the free-text clinical questions[23] and open-ended think-aloud reflection[16,17]) allows for a range of reflective experiences through which researchers, instructors, and learners can explore clinical reasoning.

This resource adds to the growing body of resources in MedEd Portal supporting the development of clinical reasoning and similar concepts (i.e., diagnostic reasoning, diagnostic decision making). For example, several current resources focus on teaching medical students explicit strategies to develop their clinical reasoning skills[24-28]. This suite of resources adds scenarios and a reflection protocols that are explicitly designed to support more experienced physicians by way of designing in increased complexity. Additionally, many of the resources currently available emphasize teaching strategies such as classroom-based case discussions[24] , case-vignettes supported by illness script worksheets[28] or case presentations of patients seen during a family medicine clerkship[25], to name a few. Others focus on strategies faculty or peers can use to assess clinical reasoning in the clinical setting[29]. Among simulation-based or SP-based scenarios, few focus explicitly on supporting everyday clinical decision making, rather, they frame clinical reasoning as an activity that supports diagnosing rarely occurring or high-risk/low frequency diagnoses[30,31]. Also, while other live scenario simulations offered in Med Ed Portal address either diabetes or angina[32-34], none offer a pairing of different cases, allowing learners to discuss the challenges brought about by the specifics of case content.
Lastly, this resource builds upon prior work in Med Ed Portal two ways: first, these scenarios take clinical reasoning skills out of the classroom or small group context and offer individual-level practice opportunities. Second, the think aloud protocol could be independently integrated with existing simulation or SP scenarios in addition to or in lieu of post-simulation debriefing. These cases and related tools offer much needed instructional material for the outpatient primary care setting (as opposed to, for instance, the emergency setting[34].

Method

This section reports on the participatory design procedures and instructional features used to develop the scenarios; the measures and reflection tools used; the procedures and logistics for scenario implementation; and the casting and training of SPs.

Participatory Design Procedures

Participatory instructional design is an approach to design that encourages the inclusion and integration of the perspectives of diverse stakeholders[35]. This approach allowed us to develop scenarios: that could be reliably implemented by the simulation lab, that represented common patient conditions, and that would support analysis of language and behavioral patterns. The scenario was developed in three phases: initial design, pilot testing, and an implementation evaluation with physician participants.

Initial design: This stage began by determining scenario goals and identifying stakeholders who could help develop scenarios to support the practice behaviors of physicians with diverse levels of expertise. Clinical stakeholders included resident and attending physicians practicing family medicine, internal medicine, and surgery. Among these individuals, most regularly taught or evaluated less experienced physicians and provided insight into common errors and practice behaviors. Simulation stakeholders included SP trainers, SPs, and operational specialists. Our stated goals were to:

a) adapt two video-based scenarios representing common patient presentations in primary care (i.e., diabetes, angina) to the live scenario-based context,

b) ensure that both adapted scenarios contained diagnostic ambiguity, which we argued would provide participants with an opportunity to consider more than a single diagnosis, and
c) embed contextual factors into one scenario to allow for more complex clinical reasoning and an opportunity to compare participant performance across these two cases.

We conducted design meetings with small groups of stakeholders to develop the scenario-based simulations (we refer the reader to the literature on Activity Theory for more information on our theoretical framework[21,36,37]). We adapted an existing context questionnaire[36] to determine what participants' *goals* or working hypotheses may be, what clinical *tools* and diagnostic *artifacts* they may request or rely upon, what clinical *guidelines* may influence their practice behaviors, and the *roles* and anticipated *activities* of other actors who may normally be present (e.g., patient care tech). In this way we developed: the simulation session workflow (Appendix A), participant expectations and instructions (Appendix B), door information (Appendices D for diabetes, I for angina), standardized patient cases (final cases in Appendices E for diabetes, J for angina) and scenario storyboards (final storyboards in Appendices G for diabetes, L for angina).

Preliminary testing: Following the initial design phase, we conducted a read-through followed by a rehearsal of each scenario. After read-throughs and discussions with two physicians, two cast SPs, and the SP educator, we further revised pertinent medical history and social and family history and identified a series of scripted key statements for each patient case (Appendices E for diabetes, J for angina). The revision history of each design change was preserved through Google documents.

Implementation evaluation: In this phase we examined scenario implementation of 12 physician participants who completed both scenarios, examining whether our design strategies resulted in physicians considering more than a single diagnosis, allowed participants to gather enough information to develop a management plan, and provided adequate complexity for interns, residents, and attendings. We conducted a content analysis of the post encounter forms (PEFs): free text questions about leading and differential diagnoses, problem list, and management decisions (Appendix N)[23]. We also analyzed the think-aloud transcriptions for the presence of reflection (in particular, reconsidering prior stances and indicating uncertainty) to better understand the broad quality of clinical reasoning. Finally, we asked participants to rate their perceptions of scenario authenticity after they completed the second think aloud-protocol (see Appendix A for workflow and Appendix R for authenticity item).

Instructional Design Features Used to Develop this Resource

We drew from Tschan and colleagues' strategy of creating scenarios that introduced an ambiguous diagnostic situation, which they define as a series of symptoms and findings which could suggest more than one diagnosis (of note, each scenario was written as a straightforward presentation of the correct diagnosis being portrayed, validated by a group of expert physicians)[14]. For example, they designed scenarios where the SP's signs and symptoms could have plausibly been attributed to anaphylaxis or tension pneumothorax, but also included information in the scenario to allow a physician participant to rule out the incorrect diagnosis[14]. We achieved this in our scenarios by incorporating a history of present illness, past medical history, and social and family history into the case where the SP presented with symptoms of the leading diagnosis (i.e., diabetes or angina) but where some of the signs and symptoms *could* also be consistent with other conditions (e.g., urinary tract infection, indigestion). We hypothesized that this diagnostic ambiguity would generate relatively complex and authentic scenarios that could be used to support the learning of physicians across their careers[6].

For one scenario (i.e., angina), in addition to diagnostic ambiguity, we introduced a contextual factor (diagnostic suggestion) to further increase complexity. Recent literature suggests that contextual factors like this may influence clinical reasoning performance, in novice and expert clinicians alike, potentially introducing significant unwanted variance (error) in patient care[10]. When contextual factors are introduced, a physician may see two patients with the same history, symptoms, and findings and yet come to two *different* diagnostic decisions[10]. We believed that the combination of ambiguity and a contextual factor in one of the cases would both be authentic and offer an opportunity to compare the two cases for relative complexity and challenge.

Selected Measures and Reflection Tools

The scenario development process described above was accompanied by a thoughtful selection of measures and reflection tools and included: an open-ended PEF eliciting clinical reasoning; a think-aloud protocol for reflection on reasoning; and a cognitive load question to check for appropriate difficulty across participants. We describe each below.

Post encounter form: To examine the clinical reasoning *process* (i.e., the steps to the diagnosis and management decisions), we used a previously published PEF that has been argued to be reliable and valid for assessing clinical reasoning (see Appendix P)[23]. This measure asks for leading and differential diagnoses, additional interview questions or exam actions participants would like to take, a problem list, supporting data for the leading diagnosis, and a management plan. We considered that this detailed open-ended measure would give us a good understanding of the process participants go through in coming to a diagnosis and treatment plan.

Think-aloud reflection: Asking someone to "think aloud" about a task, either concurrently or retrospectively, can provide insight into cognition and experience (see Appendix C for think-aloud warm-up and instructions)[16,17]. Moreover, thinking aloud has been used to great benefit in live simulation, offering a better understanding of reasoning and actions throughout the simulation[15]. Unlike some other forms of reflection (e.g., debriefing), thinking aloud involves little to no feedback during the exercise[15-17,38]. Instead, while watching the video of their own performance, participants are encouraged to provide almost a stream of consciousness reflection on their thoughts at the time of the scenario. The think-aloud literature advises the use of only minimal verbal prompting, such as "keep talking," "uh-huh," or "think aloud" if the participant pauses for more than 15-60 seconds[16,17,38]. This retrospective thinking aloud not only reveals reasoning patterns, but offers an opportunity for participants to strengthen their learning through this reflection[15].

Cognitive load: We also examined participants' cognitive load related to completing the post-encounter form using a single question provided on a separate from adapted from Brunken and colleagues[39]. We assessed participants' perceptions of their cognitive load after they completed each PEF (see Appendix A for workflow) asking participants to "please rate your invested mental effort after completing the post-encounter form" on a scale ranging from one (very low mental effort) to ten (very high mental effort). Due to the range of years of experience of participants, we included this question to check for adequate effort and engagement across participants.

Scenario Procedures and Logistics

Scheduling logistics: For each scheduled date we requested two rooms in the simulation center. The first room was used to allow participants to complete the think-aloud warm up and PEF and to re-watch their own video-recorded performance while thinking aloud. No special setup was required for the first room. The second room mimicked an outpatient clinic setting, including an exam table, a stool, a chair, a sink and a functioning headwall with an otoscope and ophthalmoscope. Participants were also provided with a stethoscope in the event they didn't bring their own. A complete supply list is included in the appendices (Appendices H for diabetes & M for angina).

Staffing requirements: We scheduled two team members, in addition to the designated SPs portraying patient roles, to support each session. The first team member was responsible for greeting each participant, ensuring that they were oriented to the simulation and think-aloud activities and that they completed all the steps of the session. The second team member was responsible for coordinating the SPs and simulation operations (e.g., giving door report, keeping time), managing the video recording and sitting with participants while they engaged in the think aloud (Appendix C). Both team members were trained to conduct the think-aloud protocol and were research associates, rather than a physician team member.

Video recording and video playback during think-alouds: To support the replaying of participant videos during their think alouds, we video recorded each scenario using two video cameras fitted with removable SD cards. In this way, one camera could act as a backup in case the primary camera failed.

Think-alouds: Following each scenario, while the participant was completing the PEF, a study team member removed the data card from the camera and inserted it into a designated computer for replaying. This same team member then read the instructions and sat with the participant during the think aloud process. During the think-aloud, the study team member was instructed to not ask questions and to limit verbal interactions to comments such as "uh huh" or "hmm" to minimize disruptions. In the event participants stopped thinking aloud for more than 15-60 seconds, the study team member gently nudged the participant by saying, "think aloud." Appendix C contains detailed warm-up and implementation instructions.

Participant procedures: On the scheduled scenario day, physician participants were oriented to the simulation rooms and the workflow of the day (Appendices A & B). They were then oriented to the think-aloud procedures they would use following the scenario (Appendix C). Instructions and practice think-aloud exercises were scripted for consistency and were implemented by study team members.

Next a study team member:

1. provided participants with the door information (Appendix D) for the first scenario and advised them to enter when ready,
2. provided participants with up to 15 minutes to complete their initial assessment, physical exam and post-assessment discussion with the SP (there was no penalty for finishing early or being stopped before completion and, depending upon time constraints, some participants were allowed to go a couple of minutes beyond 15), and
3. advised participants that the scenario would run in actual time (i.e., not sped up).

Following the scenario performance, a study team member guided participants to the designated debriefing room to:

4. complete the PEF (Appendix P),
5. complete the cognitive load question (Appendix Q)
6. review the instructions for thinking aloud (Appendix C), and
7. re-watch their own video-recorded performance while thinking aloud (this was audio recorded using a digital audio recorder [Appendix C]).

Following the first scenario, participants followed steps one through seven above for the second scenario. Participants' total time to complete these two scenarios, the related post encounter forms and think alouds, and the other informational questionnaires was approximately two hours.

Optional feedback: Because these scenarios and reflection protocols were initially used to support researching clinical reasoning processes we did not schedule time for immediate feedback. However, we recognized that participation in these scenarios could still be treated as learning experiences. Thus, following participation, we offered to scheduled time for participants to receive feedback from an attending physician on the study. These sessions were scheduled on an ad hoc basis.

Casting, Training and Quality Improvement

SP casting and training: We sought SPs similar to our designed role in age and body habitus (e.g., diabetes actress was moderately overweight). SPs were provided with the patient case (Appendices E for diabetes and J for angina) and then then rehearsed with an SP trainer as needed, drawing from a rehearsal guide (Appendices F for diabetes, K for angina). The use of a rehearsal guide was intended to support implementation fidelity because we occasionally had large breaks in time between study participants. SPs were instructed to provide information if prompted and minimize volunteering.

Quality improvement of SP performance: We developed and conducted a review of all SP portrayals to examine how consistently they implemented their roles (see Appendices N for diabetes and O for angina). This, in turn,

supported ongoing SP training needs and guided decisions about which performances were of high enough quality for analysis. For example, if an SP's performances were inconsistent with the case as written, we posited that clinical reasoning processes could be skewed. After implementation reviews, findings were shared with the SPs to improve future performance. Findings also supported ongoing scenario improvements (e.g., modifying a scripted SP response or gestural cue).

Results

Participants in this sample were 12 internal medicine, family medicine and surgery physicians, six (6) female and six (6) male. Eight (8) were resident physicians (five [5] from post graduate year 1 [PGY 1], three [3] from PGY 3) and four (4) attending. Age and gender of participants are given in Table 1.

Use of Scenario Time

For the diabetes scenario, participants' time ranged from 7:06 to 19:10 minutes (*m* = 14:38 minutes). In the stable angina scenario, participants' scenario time ranged from 11:10 to 17:15 minutes (*m* = 14:19 minutes). Two participants ran out of time and their scenarios were stopped by the study team between 17 and 19 minutes to protect participants' schedules and ensure completion of the PEF and think aloud.

Table 1. Participant Demographics (*N* =12)

Training Level	Age	Gender
Intern (PGY-1)	32	Female
	28	Male
	42	Female
	27	Female
	27	Male
Resident (PGY-3)	30	Female
	29	Male
	29	Female
Attending	55	Male
	60	Male
	38	Female
	49	Male

Differential Diagnoses and Supporting Data Listed by Participants

Diabetes: Participants considered a total of 17 independent differential diagnoses as measured by the PEF (Table 2). The most common differentials included diabetes (n = 12), hypothyroidism (n = 9), diabetes insipidus and urinary tract infection (n = 5 for the latter two). These appeared to differ by PGY: interns considered ten independent differential diagnoses, residents considered five and attendings (those having completed their initial residency) listed 12. The number of differential diagnoses listed by each participant ranged from three to six (m = 4.0). These also differed by PGY status (due to the small size of the sample, neither this nor any of the distinctions below are statistically significant): interns listed between three and six differentials (m = 4.0), residents (PGY -3) listed between three and four differentials (m = 3.3) and attendings listed between three and six differentials (m = 4.25). This range suggests that, despite the straightforwardness of the case in terms of leading diagnosis (all participants correctly listed diabetes as their leading diagnosis), there was adequate ambiguity to create other possibilities.

Content analysis of the PEF revealed that the most common supporting data participants listed included, polydipsia (*n* = 10), polyuria and fatigue (*n* = 9 for the latter two), polyphagia and recurrent yeast infections (*n* = 7 for the latter two), vision changes (*n* = 6) and obesity (*n* = 4). Participants also listed items related to past medical and family history. Among the most common were, hypertension and hypothyroid (*n* = 3) and smoking history and prior parathyroid surgery (*n* = 2). These differed by PGY status: interns listed between four and six items of supporting data (*m* = 4.8), residents listed between five and nine (*m* = 7.33) and attendings listed between four and ten *(m* = 7.5).

Table 2. Most Common Differential Diagnoses Considered for the Diabetes Scenario

Differential Diagnosis	**Frequency of listed differential diagnoses for interns (*n* = 5)**	**Frequency of listed differential diagnoses for residents (*n* = 3)**	**Frequency of listed differential diagnoses for attendings (*n* = 4)**
Type 2 Diabetes	5	3	4
Hypothyroidism	5	2	2
Diabetes Insipidus	1	3	1
Urinary Tract Infection	1	2	2
Hypercalcemia	2		
Psychogenic Polydipsia	1		1
SIADH	2		
Yeast Infection		1	1

Note: Additional diagnoses listed that received a single mention, included: anemia, bladder incontinence, glomuleronephritis, nephrotic syndrome, non-specific endocrine, non-specific autoimmune, MEN, potomania and sleep apnea.

Angina: The most common leading diagnoses were angina (*n* = 5), stable angina (*n* = 4), coronary artery disease (n = 2), and acute coronary syndrome (*n* = 1). We considered unstable angina as the correct leading diagnosis as it was the most specific, but offered near full credit for angina, angina pectoris, and stable angina.

Participants considered a total of 25 independent differential diagnoses (Table 3). The most common differentials included cardiac causes, like, coronary artery disease/acute coronary syndrome/unstable angina/stable angina (*n* = 17), followed by GERD (*n* = 9) musculoskeletal/costochondritis (*n* = 4), pulmonary embolism (*n* = 4), and peptic ulcer disease (*n* = 3). Notably, GERD was the most commonly mentioned diagnostic suggestion by SPs in the scenario. When taking into consideration PGY status, interns listed between three and nine (*m* = 4.4), residents between three and five *(m* = 3.6) and attendings between three and six (*m* = 4.75). These also appeared to differ by PGY status: interns considered 15 independent differential diagnoses, residents 10 and attendings 13.

The most common supporting data participants listed on the PEF included, chest pain (*n* = 12), which seven participants further qualified regarding onset with exertion; shortness of breath/dyspnea (*n* = 10), which six participants further qualified as also occurring with exertion, history of hypertension (*n* = 8); diabetes and smoking (*n* = 7 for the latter two), GERD (*n* = 5) and family history of cardiac disease (*n* = 3). When broken out by PGY status: interns listed between two and eight items of supporting data (*m* = 4.8), as did residents (*m* = 5) and attendings listed between two and 13 (*m* = 6.75).

Table 3. Most Common Angina Differential Diagnoses Considered

Differential Diagnosis	Frequency of listed differential diagnoses for interns (n = 5)	Frequency of listed differential diagnoses for residents (n = 3)	Frequency of listed differential diagnoses for attendings (n = 4)
Cardiac causes, like Coronary Artery Disease, Acute Coronary Syndrome, Angina/Angina Pectoris, Stable Angina, Unstable Angina[1]	8	4	5
GERD	3	2	4
Costochondritis/ Musculoskeletal Pain	2	1	1
Pulmonary Embolism		2	2
Peptic Ulcer Disease	1		2
Congestive Heart Failure	2		

Note: [1]Frequency counts exceed 12 because some participants listed more than one cardiac diagnosis. Additional diagnoses listed that received a single mention, included: anxiety, aortic dissection, arrhythmia, asthma, chronic cholelithiasis, COPD, deep vein thrombosis, enteritis, esophageal motility disorder, gastritis, myocardial infarction, non-cardiac chest pain, pancreatitis, prinzmetals angina and structural heart disease.

Management Considerations

We also examined participant PEFs for reasoning related to patient management.

Diabetes: Each suggested management, treatment, or testing option was individually scored by physician experts as correct, partially correct, or incorrect, resulting in a percentage of correct suggestions for each participant. For the diabetes case, attendings scored slightly better on the management item (*m* = 67.6%) compared to interns and residents (*m* = 56.9% for both). The most frequent lab tests requested included a blood glucose, AIC, thyroid levels/panel (*n* = 5 for each of these), a complete metabolic panel (CMP) (*n* = 4), urinalysis (*n* = 4) and urine culture (*n* = 4), and a CBC (*n* = 3). Other labs participants listed included: a urine glucose, ECG, KOH, an ABG, insulin antibodies, a cholesterol panel, a urine sodium and blood sodium. Three participants indicated they would request labs; however, they did not distinguish any specific tests.

In addition to obtaining labs, nine of twelve participants provided additional management choices that included, a) pharmacological management (e.g., use of antihyperglycemics like Metformin, an insulin trial), b) lifestyle management (e.g., nutrition, exercise) and c) referrals to other specialists (e.g., diabetes nurse educator, ophthalmologist).

Angina: For the angina case, management scores were similar, with the three residents scoring most highly (*m* = 81.9%), followed by interns (*m* = 77.8%) and attendings (*m* = 76.3%). The most frequent diagnostic test requested by participants was obtaining a stress test (*n* = 10) followed by obtaining an ECG (*n* = 9). Two participants considered requesting a chest x ray. Participants also considered obtaining additional laboratory testing, such as a CBC (*n* = 2), a CMP (*n* = 2) and cardiac enzymes (*n* = 2). Other labs mentioned included a lipid profile, A1C, and a urine glucose. Three participants indicated they would request labs; however, they did not provide further detail.

In addition to testing, four participants considered pharmaceutical management, the most common medications being a statin (*n* = 3), nitrates (*n* = 3) and aspirin (*n* = 3). Other medications listed included an ACE inhibitor, beta blockers and adjustments to the patient's current medications (i.e., HCTZ, Lisinopril). Four participants discussed whether to admit the patient or manage him in the outpatient setting and two indicated a cardiac catheterization might be necessary. These participants also prioritized administration of medications and stress testing using qualifiers, including "expedite," "ASAP" and "right away."

Think-Aloud Reflections

To explore participants' reasoning processes, we coded the think-alouds for (a) *reconsiderations*: indications that a participant would have done something differently, either in the scenario itself or the PEF and (b) *tentativeness:* words like *possibly, try, seem*, and *if* that tend to indicate uncertainty (the former were hand coded and the latter were automatically coded by Linguistic Inquiry and Word Count; LIWC)[40]. While detailed qualitative analysis is underway, we believe this initial pass offers some evidence that participants are actively reasoning by rethinking their decisions and hedging their beliefs.

Diabetes: As with PEF analysis above, we noted differences among the groups. Participants reconsidered their actions in the diabetes case between 0 and 14 times (*m* = 3.5). While significance testing was not possible (here or in any of these analyses), we noted that interns (*m* = 6.4) reconsidered actions more than residents (*m* = 2.7) who, in turn, reconsidered more than attendings (*m* = .5). Meanwhile, all participants used tentativeness markers (measured as a percentage out of the total word count), ranging from 4.5% of total words in a case to 10.1%. Interns (*m* = 7.1%) and residents (*m* = 7.5%) were more tentative in their diabetes think-alouds than attendings (*m* = 5.4%). Thus, while most participants reconsidered actions and were tentative in their phrasing to some degree in the diabetes case, attendings reconsidered less and were less tentative.

Angina: Participants reconsidered actions in the angina case between 0 and 11 times (*m* = 3.3). Interns (*m* = 4.8) and residents (*m* = 4.3) reconsidered actions more than attendings (*m* = .5). Tentativeness markers ranged from 4.2% of total words in a case to 8.9%. For the angina case, residents (*m* = 7%) were slightly more tentative than interns (*m* = 5.5%) and attendings (*m* = 5.6%). Thus, interns and residents reconsidered more actions than attendings, but residents were more tentative than either interns or attendings (again, with no statistical significance).

Cognitive Load

Diabetes: Participants' self-reported cognitive load for completing PEFs for this scenario ranged from four to eight on a scale of one to ten (*m* = 6.2; see Table 4). While the sample was too small for significance testing, we noted that PGY 1 interns found this scenario to be less complex (*m* = 5.8) than attendings (*m* = 6.8). PGY -3 residents rated it in between those groups (*m* = 6).

Angina: Participants rated the cognitive load of this scenario slightly higher than diabetes (*m* = 6.6; see Table 4), but not significantly so. Interns, residents, and attendings rated it relatively similarly (*m* = 6.6, *m* = 6.7, and *m* = 6.5 respectively).

Table 4. Self-Reported Cognitive Load by Level of Expertise (*N* =12)

Scenario	Level of Expertise	Minimum	Maximum	Mean
Diabetes	PGY 1 Interns (*n* = 5)	4	8	5.8
	PGY 2-4 Residents (*n* =3)	5	7	6
	Attendings (*n* = 4)	5	8	6.8
	Total	4	8	6.2
Angina	PGY 1 Interns (*n* = 5)	5	7	6.6
	PGY 2-4 Residents (*n* =3)	6	7	6.7
	Attendings (*n* = 4)	5	8	6.5
	Total	5	8	6.6

Participant Ratings of Scenario Authenticity

Participants rated both the diabetes and angina cases as being highly authentic, with a mean of 4.8 for diabetes and 4.6 for angina (both on a scale of one to five). While there was not enough power to test statistically, we noted that attendings rated the authenticity equal to or higher than interns or residents (see Table 5).

Table 5. Reported Scenario Authenticity by Level of Expertise (*N* =12)

Scenario	Level of Expertise	Minimum	Maximum	Mean
Diabetes	PGY 1 Interns (*n* = 5)	4	5	4.8
	PGY 2-4 Residents (*n* =3)	4	5	4.7
	Attendings (*n* = 4)	4	5	4.8
	Total	4	5	4.8
Angina	PGY 1 Interns (*n* = 5)	4	5	4.6
	PGY 2-4 Residents (*n* =3)	4	5	4.3
	Attendings (*n* = 4)	4	5	4.8
	Total	4	5	4.6

Discussion

This article describes the development and implementation of two scenarios used to formatively assess the clinical reasoning of physicians with a range of experience (i.e., interns, residents, and attendings). Findings from the implementation evaluation suggest that our strategies of including diagnostic ambiguity and contextual factors (i.e., diagnostic suggestions by the SP) may have increased complexity, possibly influencing physicians to consider a diverse range of differential diagnoses. Moreover, participants' reconsiderations, tentative language, moderate cognitive load ratings, and high authenticity ratings indicate that the design was challenging and engaging enough for interns through attendings. Of interest, we noted that while most participants selected the correct leading diagnosis, reported management choices displayed greater diversity.

These scenarios place a priority on *examining and practicing clinical reasoning behaviors.* This approach allows participants and instructors to focus not only on the outcome or solution to a diagnostic problem, but equally on the nuanced and iterative *meaning making process* leading to that solution[4,41,42]. Moreover, the inclusion of planned contextual factors provides opportunities to practice and reflect on the ways the meaning making process can shift across contexts[10]. For instance, the content analysis of the angina scenario PEF suggests that participants may have given added weight to GERD (the most frequent diagnostic suggestion) as a differential, and many participants reflected on this contextual factor in their think-alouds afterwards.

Opportunities for reflection were further supported by the think-alouds. Our brief analysis of these reflections indicates that the scenarios were complex enough for most participants, particularly newer clinicians, to reflect on possible changes to their practice through reconsiderations. Moreover, all participants used some tentativeness markers, which have been argued to indicate that an event has not been fully processed[40]. Thus, even when physicians reach a diagnosis and treatment plan, our preliminary results suggest that these cases may be complex enough to warrant some further processing.

Clinical reasoning likely differs according to level of expertise[43], as suggested by attendings' lower use of tentative language and reconsiderations compared to interns and residents. Nonetheless, the cognitive load and scenario authenticity findings reported here further support that these scenarios can provide interns, residents and attendings with a sufficiently challenging situation in which to engage. For example, two attendings had a relatively high cognitive load and the *highest* authenticity ratings when compared to residents, suggesting that these scenarios can be used across expertise levels. This approach potentially provides an alternative for those working to support the lifelong development and improvement of clinical reasoning in physicians of multiple levels of training[6,44].

Reflections on Development

Scenario-based simulation design is a complex task wherein designers attempt to plan many of the possible pathways scenario participants may take. In our experience, the result of incorporating diverse stakeholders' unique perspectives resulted in robust scenarios and being better prepared for addressing any unusual choices participants made.

For others considering participatory design approaches, we recommend that one individual be responsible for leading and coordinating the design effort, scheduling outreach to the different subject matter experts (SMEs) and supporting the occasional need to resolve conflicting team perspectives. While this coordination among multiple SMEs during the extended design and testing phases was sometimes time consuming, the process resulted in scenarios that required minimal revision during the implementation phase. This subsequently resulted in all 24 scenarios (12 diabetes and 12 angina) that we ran being of sufficient quality for inclusion in our larger study. Given the cost and scheduling constraints associated with scenarios, this added planning time seems worthwhile, minimizing the need to over-recruit study participants and preventing the disappointing loss of staff and laboratory time; funds; and participant data, time and effort.

Furthermore, incorporating diagnostic ambiguity proved to be a challenging task throughout all design phases. For example, writing detailed past, family and medical histories made it more difficult to predict which aspects participants might attend to. However, the participatory design approach made this process easier: our clinically oriented subject matter experts reviewed the SP cases multiple times to explore potential participant actions. Additionally, during the pilot phase, the SPs and SP educator highlighted the difficulty SPs might have in preparing to implement these scenarios. This helped us enhance our training and retraining strategies to include SP think alouds and the development of a rehearsal guide (Appendices F for diabetes, K for angina). During implementation the team noted the importance of tracking the variety of questions participants asked the SPs. This observation resulted in the development of the SP implementation checklist for each case (Appendices N for diabetes, O for angina). Subsequently these became an important part of our process for determining scenario implementation quality.

Through this careful design process, we were able to more consistently implement scenarios while still allowing for participant flexibility in the face of the ambiguity and contextual factors, resulting in the consideration of a variety of diagnoses and management strategies and the opportunities to reconsider these decisions.

Limitations

First, due to the difficulty recruiting participants for research, the sample size is small, only 12, making it difficult to generalize results beyond this group. Also, designing and refining these scenarios was challenging. Although our inclusion of multiple SMEs resulted in robust scenarios, taking an explicit participatory design approach was logistically challenging. For example, scheduling and coordinating meetings with SMEs required patience and there were occasional disagreements among SMEs about which aspects of the case were relevant and should be included. The lead instructional designer sought resolution through careful discussion. Additionally, during the implementation phase, we noted that a more complex scenario required more training and re-training for our SPs than initially expected. We addressed this by training SPs in pairs and providing detailed feedback using the implementation checklist (Appendices N, O). However, it should be noted that these scenarios were part of a research program, so some of these processes might be more rigorous than needed for other uses of these scenarios.

Lastly, the use of think alouds as a reflection strategy, as opposed to relying on brief faculty feedback, may be challenging for programs with time and space constraints because individual think alouds require scheduling the same amount of time as the participant's scenario and ideally a private room to complete the protocol uninterrupted. This, in fact, is one of the reasons we curtailed scenario times to approximately 15 minutes. Also, proper implementation of think alouds requires those sitting with the participant to be patient and wait until thinking aloud is complete. Most team members indicated early on that this was difficult because they often thought of questions for the participant as they listened. Yet they reported that it became easier with practice and was a valuable way to allow the participant space to reflect.

Future Directions

Developing and evaluating these scenarios highlighted the need to further examine the benefits of using scenario-based simulations for evaluating and teaching clinical reasoning, specifically focused on *management* choices. For example, the broad variation in the management choices participants considered and the effect of acuity (e.g., uncertainty about admitting or treating the angina patient in the outpatient setting) and resource availability on those plans suggest these kinds of scenarios could be important tools[45].

Additionally, since reflection is considered a vital component of simulation, the integration of open ended PEFs and think alouds could be used as a complementary reflection experience for simulation stakeholders that does not require recruiting large numbers of clinical faculty. Instead, this suite of resources is administered by trained research associates and simulation educators seeking to elicit what participants were thinking as they engaged. When used in conjunction with other simulation-based experiences relying on the support of clinical faculty or

trained debrief facilitators or standardized patients, this may offer learners a broader set of reflection experiences. Further research could be done to examine this combination of strategies.

Lastly, our strategy of integrating ambiguity did help create scenarios that were well received by diverse participants; however, because the process presented some challenges, developing systematic guidelines or a tool kit might be helpful to other simulation-based instructional designers.

References

1. Croskerry P. A Universal Model of Diagnostic Reasoning. *Academic Medicine.* 2009;84:1022-1028.
2. Heneghan C, Glasziou P, Thompson M, Bmj PR-, 2009 u. Diagnostic strategies used in primary care. *bmjcom.*
3. Elstein AS, Shulman LS, Sprafka SA. Medical Problem Solving an Analysis of Clinical Reasoning. 1978.
4. Juma S, Goldszmidt M. What physicians reason about during admission case review. *Advances in Health Sciences Education.* 2017;22:691-711.
5. Young M, Thomas A, Lubarsky S, … TB-mjo, 2018 u. Drawing Boundaries: The Difficulty in Defining Clinical Reasoning. *europepmcorg.*
6. Balogh EP, Miller BT, Ball JR. Improving Diagnosis in Health Care. 2015.
7. Norman GR, Eva KW. Diagnostic error and clinical reasoning. *Medical Education.* 2010;44:94-100.
8. Surry LT, Torre D, Durning SJ. Exploring examinee behaviours as validity evidence for multiple-choice question examinations. *Medical Education.* 2017;51:1075-1085.
9. Power A, Lemay J-F, Cooke S. Justify Your Answer: The Role of Written Think Aloud in Script Concordance Testing. *Teaching and Learning in Medicine.* 2017;29:59-67.
10. Durning SJ, Artino AR, Boulet JR, Dorrance K, van der Vleuten C, Schuwirth L. The impact of selected contextual factors on experts' clinical reasoning performance (does context impact clinical reasoning performance in experts?). *Advances in Health Sciences Education.* 2012;17:65-79.
11. Forsberg E, Ziegert K, Hult H, Fors U. Clinical reasoning in nursing, a think-aloud study using virtual patients - a base for an innovative assessment. *Nurse education today.* 2014;34:538-542.
12. Bucknall TK, Forbes H, Phillips NM, Hewitt NA, Cooper S, Bogossian F. An analysis of nursing students' decision-making in teams during simulations of acute patient deterioration. *Journal of Advanced Nursing.* 2016;72:2482-2494.
13. Prakash S, Bihari S, Need P, Sprick C, Schuwirth L. Immersive high fidelity simulation of critically ill patients to study cognitive errors: a pilot study. *BMC Medical Education.* 2017;17:36.
14. Tschan F, Semmer NK, Gurtner A, et al. Explicit Reasoning, Confirmation Bias, and Illusory Transactive Memory. *Small Group Research.* 2009;40:271-300.
15. Burbach B, Barnason S, Thompson SA. Using "Think Aloud" to Capture Clinical Reasoning during Patient Simulation. *International Journal of Nursing Education Scholarship.* 2015;12:1-7.
16. Durning SJ, Artino AR, Beckman TJ, et al. Does the think-aloud protocol reflect thinking? Exploring functional neuroimaging differences with thinking (answering multiple choice questions) versus thinking aloud. *Medical Teacher.* 2013;35:720-726.
17. Ericsson KA, Simon HA. How to Study Thinking in Everyday Life: Contrasting Think-Aloud Protocols With Descriptions and Explanations of Thinking. *Mind, Culture, and Activity.* 1998;5:178-186.
18. Alessi S. Fidelity in the design of instructional simulations. *Journal of Computer-Based Instruction.* 1988;15(2):40-47.
19. Dieckmann P, Gaba D, Rall M. Deepening the Theoretical Foundations of Patient Simulation as Social Practice. *Simulation In Healthcare: The Journal of the Society for Simulation in Healthcare.* 2007;2:183-193.
20. Kneebone RL, Scott W, Darzi A, Horrocks M. Simulation and clinical practice: strengthening the relationship. *Medical Education.* 2004;38:1095-1102.
21. Battista A. An activity theory perspective of how scenario-based simulations support learning: a descriptive analysis. *Advances in Simulation.* 2017;2:23.
22. Durning SJ, Artino AR, Pangaro LN, van der Vleuten C, Schuwirth L. Perspective: Redefining Context in the Clinical Encounter: Implications for Research and Training in Medical Education. *Academic Medicine.* 2010;85:894-901.
23. Durning SJ, Artino A, Boulet J, et al. The feasibility, reliability, and validity of a post-encounter form for evaluating clinical reasoning. *Medical Teacher.* 2012;34:30-37.
24. Weinstein A, Pinto-Powell R. Introductory Clinical Reasoning Curriculum. *MedEdPORTAL Publications.* 2016;12.
25. Felix T, Richard D, Faber F, Zimmermann J, Adams N. CoffeeTalks: An Innovative Approach to Teaching Clinical Reasoning and Information Mastery to Medical Students. *MedEdPORTAL Publications.* 2015.

26. Moroz A. Clinical Reasoning Workshop: Cervical Spine and Shoulder Disorders. *MedEdPORTAL Publications.* 2017;13.
27. Daniel M, Carney M, Khandelwal S, et al. Cognitive Debiasing Strategies: A Faculty Development Workshop for Clinical Teachers in Emergency Medicine. *MedEdPORTAL Publications.* 2017;13.
28. Levine M, Cennimo D, Chen S, Lamba S. Teaching Clinical Reasoning to Medical Students: A Case-Based Illness Script Worksheet Approach. *MedEd Portal.* 2016.
29. Weinstein A, Gupta S, Pinto-Powell R, et al. Diagnosing and Remediating Clinical Reasoning Difficulties: A Faculty Development Workshop. *MedEdPORTAL Publications.* 2017;13.
30. Beaver B, Wittler M. Toxic Ingestion/Acute Trycyclic Antidepressant (TCA) Ingestion. *MedEd Portal.* 2015.
31. Metzner J, Lombaard S, Au A, Kim S. Venous Air Embolism Curriculum. *MedEd Portal.* 2008.
32. Donegan D, Mader R, Weigel S, Kennel K. Newly Diagnosed Type 1 Diabetes Mellitus: A Resident Simulation. *MedEdPORTAL Publications.* 2013.
33. Glick S, Buchanan D, Rohr L, Kehoe L. Homeless Health Care Simulated Patient Case. *MedEdPORTAL Publications.* 2007.
34. Heitz C, Burton J, Fortuna T, Kuehl D, Perkins J, Prusakowski M. The Undifferentiated Chest Pain Patient - An Introduction to the ED Approach to the Patient. *MedEdPORTAL Publications.* 2013.
35. Könings KD, Brand-Gruwel S, Merriënboer JJG. Towards more powerful learning environments through combining the perspectives of designers, teachers, and students. *British Journal of Educational Psychology.* 2005;75:645-660.
36. Jonassen DH, Rohrer-Murphy L. Activity theory as a framework for designing constructivist learning environments. *Educational Technology Research and Development.* 1999;47:61-79.
37. Jonassen D. Revisiting activity theory as a framework for designing student-centered learning environments. *Theoretical Foundations of Learning Environments.* 2000:89-122.
38. Boren T, Ramey J. Thinking aloud: reconciling theory and practice. *IEEE Transactions on Professional Communication.* 2000;43:261-278.
39. Plass J, Moreno R, Brünken R. Measuring Cognitive Load. In: Plass JL MR, Brunken R eds, ed. *Cognitive load theory.* New York, New York: Cambridge University Press; 2010:pp 181-202.
40. Tausczik YR, Pennebaker JW. The Psychological Meaning of Words: LIWC and Computerized Text Analysis Methods. *Journal of Language and Social Psychology.* 2010;29:24-54.
41. Ilgen JS, Eva KW, Regehr G. What's in a Label? Is Diagnosis the Start or the End of Clinical Reasoning? *Journal of General Internal Medicine.* 2016;31:435-437.
42. *Understanding Clinical Reasoning from Multiple Perspectives: A Conceptual and Theoretical Overview,* 35-46(Springer, Cham 2018).
43. Schmidt HG, Mamede S. How to improve the teaching of clinical reasoning: a narrative review and a proposal. *Medical Education.* 2015;49:961-973.
44. Nestel D, Kelly M. Strategies for research in healthcare simulation. *Healthcare Simulation Education.* 2017:35-44.
45. Cook D, Sherbino J, Durning, SJ. Management reasoning: Beyond the diagnosis. *JAMA.* 2018.

Appendices

Scenario Implementation

A. Clinical Reasoning in the outpatient setting participant workflow diagram scenarios workflow diagram
B. Participant expectations and instructions script
C. Think-aloud instructions and warm up
D. Door information for diabetes
E. Diabetes standardized patient case
F. Standardized patient rehearsal guide for diabetes
G. Diabetes storyboard
H. Supplies list for diabetes
I. Door information for angina
J. Angina standardized patient case
K. Standardized patient rehearsal guide for angina
L. Angina storyboard
M. Supplies list for angina

Measures

N. Standardized patient implementation checklist for diabetes
O. Standardized patient implementation checklist for angina
P. Post encounter form (PEF)
Q. Cognitive load question
R. Scenario authenticity questionnaire

The Linguistic Effects of Context Specificity: Exploring Affect, Cognitive Processing, and Agency in Physicians' Think-Aloud Reflections

Abigail Konopasky, PhD, Steven J. Durning, MD, PhD, Anthony R. Artino, Jr., PhD, Divya Ramani, MS, and Alexis Battista, PhD

Acknowledgements: The authors would like to thank Megan Ohmer for her help in carefully transcribing the think alouds.

Author Contributions:

All authors collaborated together on the research design. Konopasky ran the analyses and wrote up the results section. Konopasky and the remaining authors co-wrote the remainder of the paper. All authors offered substantive revisions and approve of this final version of the paper.

Research Funding: This study was supported by a grant from the JPC - 1, CDMRP - Congressionally Directed Medical Research Program (# NH83382416)

Disclaimers: The views expressed in this paper are those of the authors and do not necessarily reflect the official position or policy of the US Government, Department of Defense, Department of the Navy, or the Uniformed Services University of the Health Sciences.

Abstract

Background

The literature suggests that affect, higher-level cognitive processes (e.g., critical thinking), and agency (capacity to produce an effect) are important for reasoning tasks; however, we do not know how these factors respond to context. Using situated cognition theory as a framework, and linguistic tools as a method, we explored the effects of context specificity (a physician seeing two patients with identical presentations, but coming to two different diagnoses), hypothesizing more linguistic markers of cognitive load in the presence of contextual factors (e.g., diagnostic suggestion).

Materials and Methods

In this comparative and exploratory study, 64 physicians each completed one case with contextual factors and one without. Transcribed think-aloud reflections were coded by Linguistic Inquiry and Word Count (LIWC) software for markers of affect, cognitive processes, and first-person pronouns. A MANOVA was used to inferentially compare these LIWC categories between cases with and without contextual factors. This was followed by exploratory descriptive analysis of subcategories.

Results

As hypothesized, participants used more affective and cognitive process markers in cases with contextual factors and more *I/me* pronouns in cases without. This difference reached the 95% significance level for cognitive processing words but not affective and pronominal words. Exploratory analysis revealed more negative emotions, cognitive processes of insight, and third-person pronouns in cases with contextual factors.

Conclusions

This study exposes linguistic differences arising from context specificity. These results demonstrate the value of taking a situated cognition view of a patient encounter and shows the utility of employing linguistic tools for examining clinical reasoning.

Keywords: clinical reasoning, context specificity, linguistics, emotion, agency

The Linguistic Effects of Context Specificity: Exploring Affect, Cognitive Processes, and Agency in Physicians' Think-Aloud Reflections

Introduction

Diagnostic error is a national – if not international – crisis and is frequently cited as a leading cause of death in the United States.[1,2] One important source of error relates to context specificity, a phenomenon whereby a physician sees two patients with identical presentations and yet comes to two different diagnostic decisions.[3,4] Situated cognition theory, which argues that knowing cannot be separated from context, activity, or language,[5] provides a useful framework for exploring context specificity. In the case of clinical reasoning, one way to explore context specificity is to account for patient, physician, and environmental factors. These *contextual factors* interact and, from the situated cognition perspective, clinical reasoning is constructed. Therefore, situated cognition provides a useful theoretical lens for exploring errors and other variation in physician performance that results from context specificity.

The literature suggests that affect, higher-level cognitive processes (critical thinking, problem solving and decision making[6]), and an individual's agency (a capacity to produce an effect[7,8]) are important for reasoning tasks[8–17]; however, we do not know how these factors interact and respond in different contexts. Moreover, only affect has been extensively explored in the context of *clinical* reasoning.[9–11] Linguistics provides a mechanism for understanding how different contexts may impact affect, cognitive processes, and agency, providing a potential means for better supporting physician performance in the presence of contextual factors as well as helping to unravel the vexing phenomenon of context specificity. Therefore, the purpose of this study is to use situated cognition theory and linguistic analysis to determine whether

contextual factors lead to differences in affect, cognitive processes, and individual agency, and, if so, to describe these differences.

Affect, Cognitive Processes, and Agency in Clinical Reasoning

While much of the clinical reasoning literature focuses on hypothetico-deductive aspects of cognition, emotion (an affective state characterized by arousal that results from a specific stimulus in the environment[9]) is also an integral part of the reasoning process.[9,10,18,19] The greater the magnitude (increase in level of arousal) of the emotion, the greater the possible effect on the reasoning process.[11] But some emotional arousal is present in *all* reasoning,[19] particularly in the high-stakes context of patient care, where anxiety and stress often exist.[9] Negative emotions like anxiety can cause a narrowing of attention and risk aversion, which, in turn, can increase the chance of medical error in the form of missed or delayed diagnoses.[11,20–22] Positive emotions can often support reasoning, but they can also lead to overconfidence which can, in turn, result in less information gathering during a patent encounter.[11,20,23] Recent research has also linked emotions to contextual factors, with study participants voicing primarily negative emotional reactions to various contextual factors.[24,25] One potential solution proposed by those examining diagnostic error is to increase *explicit awareness* of these emotions, taking them seriously and exploring how different contextual factors may trigger different emotional states.[10,11,26] Linguistic analysis is one such path to increased awareness of emotions,[12,27] offering a novel way of exploring how various emotions are triggered by contextual factors.

While the major *outcomes* of clinical reasoning may be diagnostic and management plans, clinical reasoning itself is also a complex *process* of meaning making that scholars are only now beginning to fully understand.[25,28–30] Contained under the umbrella of the clinical reasoning process are narrower cognitive processes like problem representation, hypothesis

generation, hypothesis testing, and metacognition (which involves, among other things, controlling and managing one's cognition in pursuit of a task[31]).[32–34] Sometimes these cognitive processes are conscious and sometimes they are unconscious,[34] and they appear to be inhibited by certain contextual factors.[24,29,35] The presence of these inhibiting contextual factors can increase cognitive load, defined as perceived mental effort.[35,36] When the cognitive load is too high for a clinician, their reasoning can be negatively affected, leading to diagnostic error.[35,37] In order to study cognitive processes in clinical reasoning and how they may be related to cognitive load, we can explore the patterns of distinct linguistic markers like *think, know*, or *consider*.[38] This allows us to examine whether and how expression of cognitive processes shifts under the influence of context specificity. Following Khawaja and colleagues, we predicted that higher cognitive load would be associated with more cognitive process markers as individuals worked to actively understand their situation.[38]

While situated and context-dependent, the process of clinical reasoning is largely *directed* by the physician. It is the physician who marshals the necessary resources – some of which may involve other people and tools – to gather information and eventually reason to a decision. Yet, as discussed above, physicians may feel uncertain or anxious in the presence of contextual factors, letting these emotions, rather than their own intentional choices, guide their reasoning.[20,21,24] We approach this through the lens of *agency*, (broadly defined as the capacity to produce an effect[7,8]), exploring whether contextual factors affect how physicians talk about themselves as agents (or not) of the reasoning process. In particular, we examined the frequency of the first-person singular pronoun *I*, since it has been argued in prior work to be indicative of a feeling of individual, intentional causation, particularly in comparison to other pronouns like generic *you*.[8,15,39] Moreover, the first-person singular pronoun has been associated with greater

depth of reflection in medical student essays[40] in one study and decreased cognitive load in team problem solving in another.[38] These studies along with the cognitive load literature suggest additional reasons why we might expect to see decreased *I*-usage in the presence of contextual factors.

In order to examine and describe potential effects of context specificity on affect, cognitive processes and individual agency so that we can better support clinicians, we pose the following research questions:

1. Does the presence of contextual factors in cases lead to differences in linguistic measures of affect, cognitive processes, or individual agency?
2. If so, what are the patterns of different subtypes of affect, cognitive processes, and agency in cases with and without contextual factors?

Based on the literature reviewed above, we hypothesized that increased cognitive load in the condition with contextual factors would lead to a greater frequency of affect and cognitive process markers and a lower frequency of first-person singular pronouns.

Materials and Methods

This study is a comparative and exploratory linguistic analysis of 128 think-aloud reflections drawn from a larger investigation[29,41,42] of context specificity and clinical reasoning at Uniformed Services University of the Health Sciences, Walter Reed National Military Medical Center, and Naval Medical Center San Diego. The study was approved by the Institutional Review Boards at all three sites. Sixty-five physicians in internal medicine, family medicine, and surgery were quasi-randomly assigned to one of two conditions: video (n = 44) or live scenario (n = 20). Participants in the video condition viewed one patient encounter with contextual factors (e.g., low English proficiency, diagnostic suggestion) and one without. After viewing each case

and determining the diagnosis (see Durning et al., 2014, for the format used[43]), participants were asked to immediately rewatch the video and "think aloud" about how they came to the diagnosis. Participants in the live scenario modality experienced the same cases, also one with contextual factors and one without, but participated in the case as a physician with a simulated participant as the patient rather than viewing a video. After giving a diagnosis in the same format as the video condition participants, they watched the encounter they had just participated in and immediately conducted a think aloud procedure. Participants in both modalities worked with cases that had typical presentations of common diseases: diabetes mellitus and unstable angina. The case content was controlled (i.e., identical presenting symptoms, language and gestures to represent those symptoms, and physical findings) so that the only differences between the cases with and without contextual factors were the contextual factors themselves.

Think-Aloud Procedure

For the think-aloud procedure, participants were asked to speak their thoughts out loud, without making judgments or offering insights, as they engaged with the task (e.g., a video of an event[44,45]). Past work has indicated that this is a reasonable measure of thinking,[45–47] as well as an effective way to assess clinical reasoning.[48–50] In this study, participants were given brief instructions and a warm-up exercise in the think-aloud method prior to engaging in the cases. Then, after either viewing the video case or participating in the live scenario and determining the diagnosis and management, they were prompted to think aloud about their thoughts leading to diagnosis and treatment. Participants were given up to 30 minutes to complete this and were allowed to stop or rewind the video.

Data Analysis

To understand how the process of clinical reasoning is affected by context specificity, we used Linguistic Inquiry and Word Count (LIWC) software. LIWC is a transparent (i.e., coded words and phrases are accessible to researchers) text analysis program that codes for affect, cognitive processes, and agency, among other psychological processes.[12] We coded all transcripts with LIWC for the broad categories of affect and cognitive processes and the subcategory of first-person singular pronouns (i.e., *I* and *me*). To control for the potential effect of varying word counts, LIWC calculates a percentage of coded categories per 100 words (e.g., if there were 10 affect-related words in a 200-word transcript, LIWC assigns that transcript a value of 5% for affect). We then conducted a repeated measures MANOVA and follow-up univariate analyses with affect, cognitive process, and first-person pronouns as the dependent variable, comparing participants' language in the cases with and without contextual factors.

To explore patterns in affect, cognition, and agency, we examined descriptive statistics of the subcategories making up affect (positive emotions, negative emotions, anxiety, anger, and sadness) and cognitive processes (insight, causal processes, certainty, tentativeness, discrepancy, and difference--described in greater detail in the results section below). We also examined descriptive statistics of other personal pronouns (*we*, *you*, *he/she*, and *they*) to better understand how individual actions interacted with the actions of others in these data.

Results

Participants were 64 internal medicine, family medicine, and surgery physicians; 22 were women and 41 were men (See Table 1 for demographic details). Think-aloud transcripts of cases without contextual factors were between 198 and 1903 words (m = 458) and those with contextual factors were between 256 and 2293 (m = 513). Across all transcripts, affective markers represented

between 1.4% and 10.4% of these words, cognitive processing words between 9.9% and 25.5%, and first-person singular pronouns between 0.2% and 9.6%.

The Effects of Context Specificity: Affect, Cognitive Processing, and Individual Agency

MANOVA results revealed significant differences in cases with and without contextual factors (Pillai's Trace = .22, $F = 5.6$, $df = [3, 61]$, $p < .01$). Follow up univariate analyses indicated that participant language contained more affective and cognitive process markers in think alouds of cases with contextual factors. Additionally, in think alouds without contextual factors, participants used more first-person singular *I/me* pronouns, suggesting a greater expression of individual agency (see Table 2). These differences were, however, statistically significant only for cognitive processing words, not for affective markers and *I/me* pronouns.

In order to better understand the differences in affect in cases with and without contextual factors, we explored the three subcategories comprising LIWC's affect category that emerged the most frequently in our data: positive emotions, negative emotions and anxiety (a subcategory of negative emotions; see Table 3). The difference between conditions resulted from more negative emotions in cases with contextual factors, where participants thought aloud about the standardized patient's emotions (stress, anxiety) and their own thought processes (e.g., thinking "that's ridiculous" about a potential diagnosis of coal worker's lung). LIWC also picked up some medical terms (e.g., *stress* test, head *trauma*, *resolves* with rest), but these uses appear in both conditions (with and without contextual factors).

Next, LIWC's cognitive processes category derives from six subcategories: insight, causal processes, certainty, tentativeness, discrepancy, and difference (see Table 4). The greatest contrast (and the only statistically significant one in our admittedly exploratory data) appears to be in terms of insight (terms associated with learning or understanding like *think, explain,*

evaluate, or *consider*[12,51]): participants talked more about their learning or understanding when contextual factors were present, more often explicitly reflecting on their *thinking* or *considering*. While the other differences were not significant, it is notable that, in the presence of contextual factors, participants seemed to use fewer markers of certainty (terms indicating a certain level of conviction like *clear*, *sure*, *certainly*, or *namely*[12,52]) and more markers of tentativeness (terms indicating a hedging or uncertain stance like *kind of*, *may*, *if*, or *anything*[12,53]). Similarly, participants made more discrepancies (terms indicating a difference between an actual and possible state like *should*, *would*, *could*, and *need*[12]) in the presence of contextual factors, often conveying a speculation about what could or would be the case, given some condition (e.g., "Her HCTZ [dose] **could** be improved....[so] her lifestyle **could** improve."). Finally, markers of causal process (terms implying that one thing gives rise to another like *how*, *based*, *because*, or *why*[12]) and difference (terms of distinction, including negation, like *but*, *really*, *not*, or *other*[12]) appeared to be similar across conditions.

Finally, examining agency beyond first-person pronouns (discussed above), participants appeared to use third-person pronouns in the presence of contextual factors to focus on the actions of *others*, often the patient or, in the video cases, the video doctor. In fact, thinking aloud about the actions of third-person singular others was the most common pronominal use across conditions (between 4.8% and 5.5% of the word count), but these exploratory analyses suggest that *he/she/him/her* usage goes up in the presence of contextual factors as *I/me* goes down. Despite the increase in cognitive processing words (which often have *I*-subjects) in the presence of contextual factors, the overall focus on the self's actions yielded to thinking aloud about the patient and video doctor, often with reference to a contextual factor. For example, here a participant reflects on the case with a patient who is not a native speaker of English: "**She**

[patient] asks **him** [doctor] about speaking Spanish and **he** says **he** only speaks English." This participant only referred to herself eight times (0.9% of words) in this case, while she referred to herself 45 times (4.3% of words) in the non-contextual factors case.

Discussion

This study demonstrates how linguistic tools can offer insight into the situated nature of the clinical reasoning process: when contextual factors are present, participants verbalize their cognitive processes more as they work to make sense of the situation and the case. Also, while not statistically significant, the trends suggested that participants voice more emotions and fewer of their own thoughts and actions (as measured by first person pronouns) in the presence of contextual factors. These findings corroborate the predictions that emerge from situated cognition and cognitive load theory that contextual factors would engender higher cognitive load and, thus, more cognitive processes and emotion and less focus on the self (versus the contextual factors themselves).

Moreover, descriptive findings from this study offer further insight into how participants react to context specificity. First, the major difference in affect markers was in negative emotions, and most of this negative affect centered on the simulated participant or the participant's own reasoning processes. While positive emotions can also affect reasoning,[11,20] negative emotions are more frequently associated with error, which is what we see in the presence of contextual factors. As with prior studies of emotion and clinical reasoning,[9,18] anxiety was common, even with physicians solving typical cases for their field. This suggests the need to be more mindful of the effects of contextual factors, including helping physicians identify and mitigate stress and anxiety during clinical encounters.

Second, the cognitive process marker that was most strongly associated with context specificity was LIWC's "insight" category, which is language associated with understanding. The presence of a contextual factor, then, appears to focus participants' verbalizations on to what they *think*, *know*, or *remember*, among other insight processes. Future work might explore how to co-opt this verbalization of insight to support deeper metacognitive practices in the presence of contextual factors.

Third, our exploration of pronouns beyond *I* indicated that the decrease in *I* pronouns was accompanied by an increase in third-person singular *he/she/it* pronouns. This suggests that the introduction of the contextual factor may be acting on clinical reasoning in part by distracting the participant away from her own reasoning actions and toward the actions of others (patient and, in the video condition, doctor). This finding further explains earlier work that found frequent mentions of contextual factors in think alouds[24,25]: the shift in focus to a contextual factor that is patient related entails a shift in focus to the patient rather than the diagnostic process *about* the patient.

Commented [1]: But in the video condition the doctor isn't them, and so they would necessarily have to use more he/she/it, right? Just because the physician isn't them (whereas in the live, it is them).

Commented [AK2R1]: But the key is that this is within subjects—so the video case people even have that shift from video case 1 to video case 2

Limitations

Our study has several important limitations. First, think alouds are not a direct measure of cognition. Instead, they are an assessment only of what individuals verbalize. Nonetheless, they provide a useful way to explore clinical reasoning and linguistic markers.[3,23,48,54] Second, LIWC is not sensitive to context, and so it sometimes miscodes certain words (e.g., "*stress* test" as affective). This linguistic "noise," however, appears to be present across both conditions, thereby allowing LIWC to detect differences. Nonetheless, future work could benefit from refinements in the linguistic software. Finally, we examine think alouds across video and live scenario

modalities. While this offers the power to discern differences, these modalities are themselves distinctly different contexts.

Practice Implications

As observed in the present study, context affects the clinical reasoning process, as predicted by situated cognition theory. Taken together with the research on errors in reasoning outcomes,[2–4] these findings argue for education around these contextual factors, perhaps through training in metacognition and awareness.[55] Moreover, the richness of these process-based measures of clinical reasoning lend themselves to a more nuanced conceptualization of "performance." These linguistic measures could be added to the growing assessment toolbox in medical education to improve early education and remediate of struggling learners. As voice recognition technology improves, automating transcription of learner reflections, LIWC could eventually be used as a formative assessment tool to alert instructors to when learners are being distracted by contextual factors and need support.

Research Implications

To begin, these findings support the value of research using situated cognition to explore context; reasoning differs in the presence of inhibiting contextual factors. Furthermore, while scholars are beginning to examine the cognitive processes and emotions inherent in physicians' clinical reasoning,[9,11,20,21,56] to our knowledge, agency has not yet been addressed. Future research could investigate, for example, whether experiences of agency shift between clinic and inpatient contexts and, if so, whether this affects clinical reasoning. Finally, these findings demonstrate the value of linguistic analysis generally and LIWC in particular. Such tools could be applied beyond the application of exploring context specificity, examining, for instance, errors present in electronic health records, assessment of diagnostic competencies, or patient-

doctor communications. If we listen carefully to what physicians say about and during the diagnostic process, we can better support them across shifting and even confusing contexts.

References

1. Donaldson MS, Corrigan JM, Kohn LT. *To Err Is Human: Building a Safer Health System*. Vol 6. National Academies Press; 2000.
2. Singh H, Graber ML. Improving Diagnosis in Health Care--The Next Imperative for Patient Safety. *N Engl J Med*. 2015;373(26):2493.
3. Durning SJ, Artino AR, Boulet JR, Dorrance K, van der Vleuten C, Schuwirth L. The impact of selected contextual factors on experts' clinical reasoning performance (does context impact clinical reasoning performance in experts?). *Adv Heal Sci Educ*. 2012;17(1):65-79.
4. Eva KW. What every teacher needs to know about clinical reasoning. *Med Educ*. 2005;39(1):98-106.
5. Brown JS, Collins A, Duguid P. Situated cognition and the culture of learning. *Educ Res*. 1989;18(1):32-42.
6. King A. Structuring peer interaction to promote high-level cognitive processing. *Theory Pract*. 2002;41(1):33-39.
7. Kaptelinin V, Nardi BA. *Acting with Technology: Activity Theory and Interaction Design*. MIT press; 2006.
8. Konopasky AW, Sheridan KM. Towards a Diagnostic Toolkit for the Language of Agency. *Mind, Cult Act*. 2016;23(2). doi:10.1080/10749039.2015.1128952
9. Kozlowski D, Hutchinson M, Hurley J, Rowley J, Sutherland J. The role of emotion in

clinical decision making: an integrative literature review. *BMC Med Educ*. 2017;17(1):255.

10. Stolper E, Van de Wiel M, Van Royen P, Van Bokhoven M, Van der Weijden T, Dinant GJ. Gut feelings as a third track in general practitioners' diagnostic reasoning. *J Gen Intern Med*. 2011;26(2):197-203.
11. Resnick ML. The effect of affect: Decision making in the emotional context of health care. In: *2012 Symposium on Human Factors and Ergonomics in Health Care*. ; 2012:39-44.
12. Tausczik YR, Pennebaker JW. The psychological meaning of words: LIWC and computerized text analysis methods. *J Lang Soc Psychol*. 2010;29(1):24-54.
13. Pennebaker JW. The Secret Life of Pronouns: What Our Words Say About Us (NY: Bloomsbury). 2011.
14. Pennebaker JW, Graybeal A. Patterns of natural language use: Disclosure, personality, and social integration. *Curr Dir Psychol Sci*. 2001;10(3):90-93.
15. Konopasky A, Ramani D, Ohmer M, Durning SJ, Artino Jr AR, Battista A. Why health professions education needs functional linguistics: the power of 'stealth words.' *Med Educ*. 2019.
16. Bandura A. Toward a psychology of human agency. *Perspect Psychol Sci*. 2006;1(2):164-180.
17. Weiner B. An attributional theory of achievement motivation and emotion. *Psychol Rev*. 1985;92(4):548.
18. Simpkin AL, Murphy Z, Armstrong KA. A randomized experimental study to assess the effect of language on medical students' anxiety due to uncertainty. *Diagnosis*. 2019.

19. Damasio AR. *Descartes' Error*. Random House; 2006.
20. Heyhoe J, Birks Y, Harrison R, O'Hara JK, Cracknell A, Lawton R. The role of emotion in patient safety: are we brave enough to scratch beneath the surface? *J R Soc Med*. 2016;109(2):52-58.
21. Surry LT, Torre D, Trowbridge RL, Durning SJ. A mixed-methods exploration of cognitive dispositions to respond and clinical reasoning errors with multiple choice questions. *BMC Med Educ*. 2018;18(1):277.
22. Mamede S, Van Gog T, Schuit SCE, et al. Why patients' disruptive behaviours impair diagnostic reasoning: a randomised experiment. *BMJ Qual Saf*. 2017;26(1):13-18.
23. Estrada CA, Isen AM, Young MJ. Positive affect facilitates integration of information and decreases anchoring in reasoning among physicians. *Organ Behav Hum Decis Process*. 1997;72(1):117-135.
24. McBee E, Ratcliffe T, Picho K, et al. Consequences of contextual factors on clinical reasoning in resident physicians. *Adv Heal Sci Educ*. 2015;20(5):1225-1236.
25. Durning S, Artino Jr AR, Pangaro L, van der Vleuten CPM, Schuwirth L. Context and clinical reasoning: understanding the perspective of the expert's voice. *Med Educ*. 2011;45(9):927-938.
26. Trowbridge RL. Twelve tips for teaching avoidance of diagnostic errors. *Med Teach*. 2008;30(5):496-500.
27. Kahn JH, Tobin RM, Massey AE, Anderson JA. Measuring emotional expression with the Linguistic Inquiry and Word Count. *Am J Psychol*. 2007:263-286.
28. Ilgen JS, Eva KW, Regehr G. What's in a Label? Is Diagnosis the Start or the End of Clinical Reasoning? *J Gen Intern Med*. 2016;31(4):435-437. doi:10.1007/s11606-016-

3592-7

29. Konopasky AW, Ramani D, Ohmer M, et al. It Totally Possibly Could Be: How A Group Of Military Physicians Reflect On Their Clinical Reasoning In The Presence Of Contextual Factors. *Mil Med*.
30. McBee E, Ratcliffe T, Picho K, et al. Contextual factors and clinical reasoning: differences in diagnostic and therapeutic reasoning in board certified versus resident physicians. *BMC Med Educ*. 2017;17(1):211.
31. Cleary TJ, Durning S, Gruppen L, Hemmer P, Artino A. Self-regulated learning in medical education. *Oxford Textb Med Educ*. 2013:465-477.
32. Charlin B, Lubarsky S, Millette B, et al. Clinical reasoning processes: unravelling complexity through graphical representation. *Med Educ*. 2012;46(5):454-463.
33. Patel V, Arocha J, Zhang J. Thinking and reasoning in medicine. In: Holyoak K, Morrison R, eds. *The Cambridge Handbook of Thinking and Reasoning*. Cambridge University Press; 2005:727-750.
34. ten Cate O, Durning SJ. Understanding Clinical Reasoning from Multiple Perspectives: A Conceptual and Theoretical Overview. 2018:35-46. doi:10.1007/978-3-319-64828-6_3
35. Ratcliffe TA, McBee E, Schuwirth L, et al. Exploring Implications of Context Specificity and Cognitive Load in Residents. *MedEdPublish*. 2017;6.
36. Paas F, Tuovinen JE, Tabbers H, Van Gerven PWM. Cognitive load measurement as a means to advance cognitive load theory. *Educ Psychol*. 2003;38(1):63-71.
37. Fraser K, Ma I, Teteris E, Baxter H, Wright B, McLaughlin K. Emotion, cognitive load and learning outcomes during simulation training. *Med Educ*. 2012;46(11):1055-1062.
38. Khawaja MA, Chen F, Marcus N. Measuring cognitive load using linguistic features:

implications for usability evaluation and adaptive interaction design. *Int J Hum Comput Interact*. 2014;30(5):343-368.

39. Arnold J, Clarke DJ. What is 'agency'? Perspectives in science education research. *Int J Sci Educ*. 2014;36(5):735-754.
40. Howe A, Barrett A, Leinster S. How medical students demonstrate their professionalism when reflecting on experience. *Med Educ*. 2009;43(10):942-951.
41. Ohmer M, Durning SJ, Kucera W, et al. Clinical Reasoning in the Ward Setting: A Rapid Response Scenario for Residents and Attendings. *MedEdPORTAL*. 2019;15.
42. Battista A, Konopasky A, Ramani D, et al. Clinical reasoning in the primary care setting: Two scenario-based simulations for residents and attendings. *MedEdPORTAL*. 2018;14(10773).
43. Durning SJ, Artino A, Boulet J, et al. The feasibility, reliability, and validity of a post-encounter form for evaluating clinical reasoning. *Med Teach*. 2012;34(1):30-37. doi:10.3109/0142159X.2011.590557
44. Ericsson KA, Krampe RT, Tesch-Römer C. The role of deliberate practice in the acquisition of expert performance. *Psychol Rev*. 1993;100(3):363.
45. Ericsson KA, Simon HA. How to study thinking in everyday life: Contrasting think-aloud protocols with descriptions and explanations of thinking. *Mind, Cult Act*. 1998;5(3):178-186.
46. Durning SJ, Artino AR, Beckman TJ, et al. Does the think-aloud protocol reflect thinking? Exploring functional neuroimaging differences with thinking (answering multiple choice questions) versus thinking aloud. *Med Teach*. 2013;35(9):720-726. doi:10.3109/0142159X.2013.801938

47. Fox MC, Ericsson KA, Best R. Do procedures for verbal reporting of thinking have to be reactive? A meta-analysis and recommendations for best reporting methods. *Psychol Bull*. 2011;137(2):316.
48. Burbach B, Barnason S, Thompson SA. Using "think aloud" to capture clinical reasoning during patient simulation. *Int J Nurs Educ Scholarsh*. 2015;12(1):1-7.
49. Durning SJ, Graner J, Artino Jr AR, et al. Using functional neuroimaging combined with a think-aloud protocol to explore clinical reasoning expertise in internal medicine. *Mil Med*. 2012;177(suppl_9):72-78.
50. Funkesson KH, Anbäcken E-M, Ek A-C. Nurses' reasoning process during care planning taking pressure ulcer prevention as an example. A think-aloud study. *Int J Nurs Stud*. 2007;44(7):1109-1119.
51. Shaw BR, Hawkins R, McTavish F, Pingree S, Gustafson DH. Effects of insightful disclosure within computer mediated support groups on women with breast cancer. *Health Commun*. 2006;19(2):133-142.
52. Moons WG, Spoor JR, Kalomiris AE, Rizk MK. Certainty broadcasts risk preferences: Verbal and nonverbal cues to risk-taking. *J Nonverbal Behav*. 2013;37(2):79-89.
53. Joksimovic S, Gasevic D, Kovanovic V, Adesope O, Hatala M. Psychological characteristics in cognitive presence of communities of inquiry: A linguistic analysis of online discussions. *internet High Educ*. 2014;22:1-10.
54. Oliver EJ, Markland D, Hardy J, Petherick CM. The effects of autonomy-supportive versus controlling environments on self-talk. *Motiv Emot*. 2008;32(3):200-212.
55. Trowbridge RL, Dhaliwal G, Cosby KS. Educational agenda for diagnostic error reduction. *BMJ Qual Saf*. 2013;22(Suppl 2):ii28-ii32.

56. McBee E, Ratcliffe T, Goldszmidt M, et al. Clinical reasoning tasks and resident physicians: what do they reason about? *Acad Med*. 2016;91(7):1022-1028.

Table 1
Demographic variables arranged by study condition

	Video Condition		Live Scenario Condition	
	Mean (*SD*)	Range	Mean (*SD*)	Range
Age in years	35 (8.8)	26-67	37 (10.3)	26-61
Years of experience	7 (8.6)	0-39	8 (10.7)	0-35

Table 2
Univariate tests of LIWC affect, cognitive process, and individual agency markers in cases with and without contextual factors

	No Contextual Factors M (SD)	Contextual Factors M (SD)	*F*-test, significance
Affect	4.1% (1.68)	4.5% (1.27)	$F = 4.8$, $p = .06$
Cognitive Processes	16.7% (3.57)	17.5% (2.67)	$F = 4.1$, $p < .05$
First-person singular *I/me* pronouns	3.8% (2.26)	3.5% (2)	$F = 3.9$, $p = .05$

Table 3
Results of exploratory analyses of LIWC affect category markers in cases with and without contextual factors

	No Contextual Factors		Contextual Factors		*T*-test (effect size)[1]	Example
	M (SD)	Range	M (SD)	Range		
Positive emotions	2.3% (1)	0.7%-6.8%	2.4% (1.27)	0.5%-8%	$t = .5$, $p = .6$ ($d = .09$)	It [blood pressure] looks **pretty good**.
Negative emotions	**1.7%** (0.84)	0.3%-4.4%	**2.1%** (0.83)	0.6%-4.2%	$t = 3.1$, $p = .003$ ($d = .48$)	This [doctor-patient exchange] is very **awkward**.
Anxiety[2]	**0.5%** (0.41)	0-1.7%	**0.7%** (0.4)	0-1.8%	$t = 3.5$, $p = .001$ ($d = .49$)	[Patient works] **stressful** long hours. He could have **anxiety.**

[1]These exploratory analyses are not inferential; these statistics are provided for ease of interpretation.
[2]This is a sub-category of negative emotions.

Table 4
Results of exploratory analyses of LIWC cognitive process category markers in cases with and without contextual factors

	No Contextual Factors		Contextual Factors		*T*-test (effect size)[1]	Example
	M (SD)	Range	M (SD)	Range		
Insight	**3.2%** (1.44)	0.3%- 8.2%	**3.6%** (1.06)	1.3%- 6.2%	$t = 2.1$, $p = .04$ ($d = .32$)	**Realized** I'd forgotten to ask about smoking.
Causal processes	1.7% (0.81)	0-3.8%	1.7% (0.65)	0.3%- 3%	$t = .28$, $p = .78$ ($d = .04$)	That is probably not unstable [angina] **because** it's not worsening
Certainty	1.2% (.056)	0-3.2%	1.1% (0.57)	0- 2.6%	$t = .44$, $p = .67$ ($d = 1.77$)	I make **sure** I'm not missing anything.
Tentativeness	5.8% (1.66)	2.6%- 10.6%	6.1% (1.68)	2.9%- 10.5%	$t = .98$, $p = .33$ ($d = .18$)	The patient **seems** uncomfortable.
Discrepancy	2.4% (1.25)	0.4%- 6.9%	2.6% (0.94)	0.9%- 4.4%	$t = 1.07$ $p = .29$ ($d = .18$)	An infection like the flu **could** be a trigger.
Difference	4.9% (1.17)	1.3%- 7.5%	5% (1.37)	2.3%- 9.3%	$t = 1.03$, $p = .31$ ($d = .08$)	She has a slightly elevated pulse, but **not** tachycardic.

[1]These exploratory analyses are not inferential; these statistics are provided for ease of interpretation.

Table 5
Results of exploratory analyses of LIWC personal pronoun markers in cases with and without contextual factors

	No Contextual Factors		Contextual Factors		*T*-test (effect size)[1]	Example
	M (SD)	Range	M (SD)	Range		
First person singular (*I/me*)	3.8% (2.26)		3.5% (2)		$t = 1.97$, $p = .05$ ($d = .14$)	**I'**d want more information from the lungs.
First person plural (*we*)	0.3% (0.29)		0.3% (0.3)		$t = 0.19$, $p = .85$ ($d = .02$)	**We'**re seeing that the polydipsia and fatigue has been progressive.
Second person or generic (*you*)	0.8% (0.78)		0.8% (0.8)		$t = 0.48$, $p = .63$ ($d = .05$)	Not sure why **you**'re asking about alcohol. [reference to doctor in video]
						You do see maybe a little flattening of the diaphragms. [reference to physicians generally]
Third person singular (*he/she/him/her*)	**4.8%** (1.97)		**5.5%** (2.16)		$t = 2.5$, $p = .02$ ($d = .32$)	**He** mentioned that **he** felt like it was GERD.
Third person plural (*they)*	0.2% (0.32)		0.2% (0.26)		$t = 0.97$, $p = .34$ ($d = .12$)	I'm really curious what **they** [patient generally] think and what **they**'re worried about.

[1]These exploratory analyses are not inferential; these statistics are provided for ease of interpretation.

Awareness and Reflection: The Results of an Intervention to Address Context Specificity

Introduction

Clinical reasoning is fundamental to every physician in practice.[1] Clinical reasoning involves a number of activities such as information gathering, formulating a differential diagnosis, providing diagnostic justification and making diagnostic and therapeutic plans.[2–4] Mistakes in clinical reasoning undoubtedly contribute significantly to diagnostic errors, which are hypothesized to account for approximately 15% of the errors in primary care.[5] There is mounting evidence that these errors cannot solely be attributed to gaps in physician knowledge or training.[6,7] Indeed, recent research has identified the phenomenon of *context specificity*, defined as a physician seeing two identical patient presentations from a content perspective (e.g. identical histories, physical exams, labs and the same diagnosis) yet arriving at two different diagnostic decisions.[6,8] In other words, in these situations, something other than the content is driving the physician's decisions leading to unwanted variation in physician performance.

The primary theoretical framework for this article is situated cognition. From this theoretical perspective clinical reasoning is situated (or located) within the specifics of the encounter (e.g. the patient, the physician and the environment) and clinical reasoning processes and outcomes dynamically emerge from and are shaped by these specifics of the situation.[9] In situated cognition there are a host of interactions between individuals and their physical, social, and cultural systems. From this perspective, context specificity is associated with *contextual factors*, elements arising from patients, physicians, clinical environment and the interactions among all three.[6] When these contextual factors are *distracting* (e.g., a patient suggesting an incorrect diagnosis, an electronic health record not functioning optimally, compressed time for

completing an encounter), this can negatively affect clinical reasoning performance and ultimately lead to diagnostic error.[6,8]

One proposed strategy for reducing diagnostic error is the use of reflective practice.[10–12] In reflective practice, there are a number of behaviors and reasoning processes that occur in response to complex clinical problems. These include 1) a search for alternative explanations of the problem, 2) exploration of the consequences of such alternative explanations that leads to predictions to be tested by the acquisition of new data, 3) a testing of these predictions against data and a reframing of the problem, and 4) a critical review of one's own assumptions and conclusions about the problem (meta-reasoning).[10] Part of reflective practice for the reduction of error can also be related to metacognition, stepping back from the immediate situation to reflect on the thinking process.[13] However, these educational strategies have not yet been employed to determine if they can mitigate the negative effects of context specificity in an authentic (e.g. simulated) environment. This may enhance physicians' ability to recognize and possibly reduce the adverse impact of distracting contextual factors on clinical reasoning performance and reduce diagnostic errors. The development and exploration of effective educational strategies that address specific aspects of the clinical context are important for a future implementation of new instructional tools in the training of medical students, residents and physicians.

An optimal way to test out these strategies is through a simulated scenario, which creates a complex, highly interactive context. This can help us to disentangle the potential effects of distracting contextual factors on clinical reasoning and error and assess the outcome of potential educational interventions designed to ameliorate context specificity.[14,15] Scenario-based simulations provide an environment that is similar to an authentic clinical setting, allowing for different degrees of contextual and cognitive complexity in the clinical encounter while

controlling for specific leading and differential diagnoses.[15–17] Further, there is evidence that scenario-based simulations provide effective environments for assessing clinicians' performance in activities associated with clinical reasoning.[16,18]

The purpose of this study was to assess clinical reasoning performance during a simulated encounter, comparing physicians who received an educational intervention (a computer-based tutorial on contextual factors and their potential link to diagnostic errors and a think-aloud exercise) with those who did not. Our hypothesis was that the use of this dual intervention—tutorial and think-aloud—would enhance clinical reasoning performance measured by a post encounter form (PEF), leading to better performance by intervention participants.

Materials and Methods

This comparative experimental study examined whether an intervention (consisting of an interactive computer training and a think-aloud reflection) designed to support physicians' clinical reasoning in the presence of distracting contextual factors improved clinical reasoning performance, as measured by a PEF, an open-ended series of diagnostic and management questions for which we have previously gathered validity evidence.[6,19,20]

Population

Practicing military physicians in internal medicine, family medicine, and surgery from the Uniformed Services University for the Health Sciences, Walter Reed National Military Medical Center, and the University of Texas Health Science Center were assigned to either the intervention (n =20) or control (n =19) condition, based on scheduling availability. Institutional Review Boards at all three sites approved this research (complying with World Medical Association Declaration of Helsinki).

Study Design

The study included two groups of participants. The control group began with the simulated encounter, followed by the PEF (the outcome measure for this study) which closely mirrors typical practice.[19] The intervention group participated in an interactive computer-based training module, completed a simulated encounter, did the think-aloud reflection, and then completed the PEF. (Control participants also completed the think-aloud exercise, but only after completion of the outcome measure; see Figure 1)

Data Collection

Training module. Participants in the intervention condition completed a computer-based clinical reasoning and diagnostic error training module covering: the nature of clinical reasoning and diagnostic error, the role of context specificity and contextual factors in diagnostic error, and the presentation of a reflection strategy (thinking aloud) for countering the potentially harmful effects of distracting contextual factors. Each of the three sections was accompanied by open-ended written reflection questions connecting the topics to participants' own practice and experience. This module took approximately 20 minutes. See Table 1 for training details.

Simulated encounter. Participants in both groups engaged in the identical standardized patient encounter. This encounter was designed in three phases, using a participatory design procedure involving clinical and simulation stakeholders (see Battista et al., 2018 for a description of the design process and all supporting materials[21]). A male standardized patient (SP) was trained to portray someone with unstable angina and a distracting contextual factor (patient conveys diagnostic suggestion of GERD). Participants were given details about the patient's initial complaint (chest pain) and then entered a room designed much like the outpatient clinic rooms where they practice, where they introduced themselves to the SP and conducted a

history and physical. Encounters were stopped if they reached 18 minutes, but no time penalties were given. Encounter time ranged from 11 to 18 minutes.

Think aloud. Immediately after the encounter, intervention participants were reminded of the think-aloud practice they had done at the end of the computer-based training and were given additional practice opportunity if needed. When ready, they were asked to watch the video tape of their encounter while thinking aloud, without judgment or analysis, about what was going through their heads as they came to the diagnosis and treatment. Previous research has argued that think aloud exercises provide a useful window into cognition and experience.[22–24] Additionally, recent work on how participants reconsider their thought processes during think aloud exercises suggests that they may also be a valuable tool for reflection in the presence of contextual factors (Konopasky et al., unpublished manuscript).

Outcome Instrument (PEF). Clinical reasoning performance was captured through the PEF, a measure developed and used in prior work for which we have gathered considerable validity evidence.[6,19,20] It asks participants for: 1) additional information they would like to obtain by history, 2) additional physical exam actions they would take, 3) a problem list, 4) a differential diagnosis, 5) a leading diagnosis, 6) supporting evidence for the diagnosis, and 7) management plans. For this analysis we focused on steps (4-7) because we were interested in exploring the cognitive processes of specific elements of clinical reasoning such as differential diagnosis, supportive evidence, and diagnostic and therapeutic plans. Participants were given up to 30 minutes (determined to be ample time in prior studies[6,21]) to complete the PEF. We used a scoring key developed by a panel of board-certified internists in prior research[6] that assigns each free-text response a point value (participants gave multiple responses for most items) for correct (2 points), partially correct (1 point), or incorrect (0 points). All responses were scored by at least

two raters, coming to consensus on disagreements and updating the scoring sheet with the decision. In order to compare across participants, who gave varying numbers of responses for each item, we created percentage scores for each item, dividing the total number of points by the total number of possible points (e.g., someone who offered five pieces of evidence for the leading diagnosis had a total possible score of 10 for that item).

Analysis

First, descriptive statistics were calculated separately for each group (intervention and control). Then, to determine if there were differences in clinical reasoning performance across our four chosen variables between groups while also controlling for increased risk of Type I error that occurs with multiple univariate tests, we conducted a multivariate analysis of covariance (MANCOVA), with percentage scores on the four PEF items as dependent variables and group (intervention versus control) as the independent variable. Age and number of years since graduation from medical school were used as covariates to control for years of experience. A power analysis indicated that a total sample size of 51 was needed to detect a medium-sized effect at a significance level of .05, so the study was underpowered with $n = 39$.

Results

Participants in the control and intervention groups were equally distributed in terms of gender (10 females out of 19 for control; 10 out of 20 for intervention), but in the control group the range of ages was wider and they had more years of experience (i.e., control participants were older and had been practicing longer; see Table 2 for demographic details).

Means for the four outcome variables ranged from 65% to 80% for control group participants and 70% to 87% for intervention group participants. See Table 3 for means and ranges of outcome variables by group.

MANCOVA results revealed no significant differences between intervention and control groups (Pillai's Trace = .20, $F = 1.9$, $df = [4, 29]$, $p = .15$, $\eta_p^2 = .2$; See Table 4 for details). The Box *M* test did not indicate a significant violation of the assumption of homogeneity of variances ($F = 1.8$, $p = .06$). Levene's test, however, only indicated equal variances for supporting evidence; the other three variables violated the assumption.

Discussion

We developed an intervention to attempt to mitigate the negative effects of context specificity on clinical reasoning performance that included a computer-based training module to raise awareness of these effects and a think-aloud exercise on participation in a simulated clinical encounter. We hypothesized that this computer-based training and think-aloud exercise would lessen the negative impact of distracting contextual factors on diagnostic and therapeutic accuracy of participants in the intervention group compared to the control group. However, we found no statistically significant difference between the two groups. Nonetheless, we believe this work represents an important contribution to the scholarship on the effects awareness and reflection may have on clinical reasoning performance and patient care, as discussed below.[25,26]

Situated cognition anticipates the potentially negative effects of distracting contextual factors, since an individual's clinical reasoning is interconnected with the social and environmental elements of a patient encounter.[9,27] Yet the environment created by distracting contextual factors also creates opportunities; Ng and colleagues refer to "indeterminate zones of practice—uncertain, unstable, unique or value-conflicted practice situations" that provide opportunities for the development of clinical practice (p. 463).[28] Contextual factors may move clinicians into these zones. Explicit education around and reflective practice upon distracting contextual factors in these zones may help clinicians develop better *situation awareness* to

trigger reflection in future encounters with contextual factors.[29] Explicit discussion of biases and "diagnostic timeouts" to reflect on these biases have been suggested more broadly as error reducing techniques[30] and this intervention builds on those insights.

Additionally, this intervention offers a somewhat different approach to reflective practice from other scholars,[10–12,26] asking participants to "think aloud" without explanation, judgment, or structure.[22,24] While this approach may sacrifice some of the benefits of specific reasoning instructions,[26] it may offer a more comfortable space for reflective practice than other, more directed approaches. Boud and Walker point to the importance of the "micro-contexts" instructors create for reflection (defined slightly differently by these authors, but still relevant to our context), particularly in terms of instructors' potential power over the learner.[31] By removing the instructor from the room altogether and with minimal interruption (beyond cuing learners to continue when they fall silent), think aloud exercises offer an opportunity to reflect across a full encounter that (in concert with other reflective practices) "permit[s] the making of meaning" (p. 10).[31] Even if this act of meaning-making does not change the decision made in a particular case, it may serve to "promote adaptive expertise and practical wisdom" (p. 1048) as physicians develop over the course of a career.[32]

There were several limitations to this study. First the sample size did not give us enough power to adequately determine group differences. The intervention group was younger (in fact, the age difference between groups was statistically significant) and less experienced than the control group, which may have dampened the impact of the intervention as experience would be expected to improve results.[33] Future studies should include a greater number of participants with similar or at least equivalent baseline characteristics. Second our intervention was comprised of two parts, the computer-based training and the think-aloud exercise, separated by the encounter;

the sequence of these two interventions may have affected the overall impact while hampering our ability to disentangle individual effects. Finally, while our PEF has validity evidence for its use,[6,19–21] it is one tool and as such provides a limited view on clinical reasoning performance.

In conclusion, while we found no statistically significant differences between the two groups, this study suggests the potential utility of an intervention providing (a) education and awareness of contextual factors and (b) a "micro-context" for safe reflective practice. Moreover, guided by situated cognition, this intervention asks participants to develop situation awareness that moves beyond the self's actions to the broader range of interactions and systemic influences that comprise a clinical encounter. Future work could explore what cues trigger this kind of awareness in physicians and how sensitivity to those cues shifts over time, helping us to more fully understand not only context specificity, but also the adaptive expertise that characterizes truly excellent physicians.

Author Contributions:

All authors collaborated together on the research design and data collection. Torre, Haynes, Woodard, and Durning did the coding, coming to consensus. Konopasky ran the analyses and wrote up the results section. Konopasky and the remaining authors co-wrote the remainder of the paper. All authors offered substantive revisions and approve of this final version of the paper.

Research Funding: This study was supported by a grant from the JPC - 1, CDMRP - Congressionally Directed Medical Research Program (#NH83382416)

Competing Interest: The funding organization played no role in the study design; in the collection, analysis, and interpretation of data; in the writing of the report; or in the decision to submit the report for publication.

Disclaimers: The views expressed in this paper are those of the authors and do not necessarily reflect the official position or policy of the US Government, Department of Defense, Department of the Navy, or the Uniformed Services University of the Health Sciences.

References

1. Norman G. Research in clinical reasoning: past history and current trends. *Med Educ*. 2005;39(4):418-427. doi:10.1111/j.1365-2929.2005.02127.x
2. Elstein AS, Shulman LS, Sprafka SA. Medical problem solving an analysis of clinical reasoning. 1978.
3. Croskerry P. A universal model of diagnostic reasoning. *Acad Med*. 2009;84(8):1022-1028.
4. Young M, Thomas A, Lubarsky S, et al. Drawing boundaries: the difficulty in defining clinical reasoning. *Acad Med*. 2018;93(7):990-995.
5. Graber ML. The incidence of diagnostic error in medicine. *BMJ Qual Saf*. 2013;22(Suppl 2):ii21-ii27.
6. Durning SJ, Artino AR, Boulet JR, Dorrance K, van der Vleuten C, Schuwirth L. The impact of selected contextual factors on experts' clinical reasoning performance (does context impact clinical reasoning performance in experts?). *Adv Heal Sci Educ*. 2012;17(1):65-79.
7. Eva KW. What every teacher needs to know about clinical reasoning. *Med Educ*. 2005;39(1):98-106.
8. Ratcliffe TA, McBee E, Schuwirth L, et al. Exploring Implications of Context Specificity and Cognitive Load in Residents. *MedEdPublish*. 2017;6.
9. Durning SJ, Artino AR. Situativity theory: a perspective on how participants and the environment can interact: AMEE Guide no. 52. *Med Teach*. 2011;33(3):188-199.
10. Mamede S, Schmidt HG, Penaforte JC. Effects of reflective practice on the accuracy of medical diagnoses. *Med Educ*. 2008;42(5):468-475.

11. Mamede S, Schmidt HG. The structure of reflective practice in medicine. *Med Educ*. 2004;38(12):1302-1308.

12. Mamede S, Schmidt HG, Rikers R. Diagnostic errors and reflective practice in medicine. *J Eval Clin Pract*. 2007;13(1):138-145.

13. Croskerry P. The importance of cognitive errors in diagnosis and strategies to minimize them. *Acad Med*. 2003;78(8):775-780.

14. Kneebone RL, Scott W, Darzi A, Horrocks M. Simulation and clinical practice: strengthening the relationship. *Med Educ*. 2004;38(10):1095-1102.

15. Battista A. An activity theory perspective of how scenario-based simulations support learning: a descriptive analysis. *Adv Simul*. 2017;2(1):23.

16. Kneebone RL, Kidd J, Nestel D, et al. Blurring the boundaries: scenario-based simulation in a clinical setting. *Med Educ*. 2005;39(6):580-587.

17. Dieckmann P, Gaba D, Rall M. Deepening the Theoretical Foundations of Patient Simulation as Social Practice. *Simul Healthc J Soc Simul Healthc*. 2007;2(3):183-193. doi:10.1097/SIH.0b013e3180f637f5

18. Daniel M, Rencic J, Durning SJ, et al. Clinical reasoning assessment methods: a scoping review and practical guidance. *Acad Med*. 2019;94(6):902-912.

19. Durning SJ, Artino A, Boulet J, et al. The feasibility, reliability, and validity of a post-encounter form for evaluating clinical reasoning. *Med Teach*. 2012;34(1):30-37. doi:10.3109/0142159X.2011.590557

20. McBee E, Ratcliffe T, Picho K, et al. Contextual factors and clinical reasoning: differences in diagnostic and therapeutic reasoning in board certified versus resident physicians. *BMC Med Educ*. 2017;17(1):211.

21. Battista A, Konopasky A, Ramani D, et al. Clinical reasoning in the primary care setting: Two scenario-based simulations for residents and attendings. *MedEdPORTAL*. 2018;14(10773).

22. Durning SJ, Artino AR, Beckman TJ, et al. Does the think-aloud protocol reflect thinking? Exploring functional neuroimaging differences with thinking (answering multiple choice questions) versus thinking aloud. *Med Teach*. 2013;35(9):720-726. doi:10.3109/0142159X.2013.801938

23. Burbach B, Barnason S, Thompson SA. Using "think aloud" to capture clinical reasoning during patient simulation. *Int J Nurs Educ Scholarsh*. 2015;12(1):1-7.

24. Ericsson KA, Simon HA. How to study thinking in everyday life: Contrasting think-aloud protocols with descriptions and explanations of thinking. *Mind, Cult Act*. 1998;5(3):178-186.

25. Mann K, Gordon J, MacLeod A. Reflection and reflective practice in health professions education: a systematic review. *Adv Heal Sci Educ*. 2009;14(4):595.

26. Mamede S, Schmidt HG. Reflection in medical diagnosis: a literature review. *Heal Prof Educ*. 2017;3(1):15-25.

27. Konopasky AW, Ramani D, Ohmer M, et al. It Totally Possibly Could Be: How A Group Of Military Physicians Reflect On Their Clinical Reasoning In The Presence Of Contextual Factors. *Mil Med*.

28. Ng SL, Kinsella EA, Friesen F, Hodges B. Reclaiming a theoretical orientation to reflection in medical education research: a critical narrative review. *Med Educ*. 2015;49(5):461-475.

29. Moulton CA, Regehr G, Mylopoulos M, MacRae HM. Slowing down when you should: a

new model of expert judgment. *Acad Med*. 2007;82(10):S109-S116.

30. Trowbridge RL. Twelve tips for teaching avoidance of diagnostic errors. *Med Teach*. 2008;30(5):496-500.

31. Boud D, Walker D. Promoting reflection in professional courses: The challenge of context. *Stud High Educ*. 1998;23(2):191-206.

32. Epstein RM. Reflection, perception and the acquisition of wisdom. *Med Educ*. 2008;42(11):1048-1050.

33. Ericsson KA. The influence of experience and deliberate practice on the development of superior expert performance. *Cambridge Handb Expert Expert Perform*. 2006;38:685-705.

Understanding Context Specificity: The Effect of Contextual Factors on Clinical Reasoning

Introduction

Diagnostic error is a problem at the forefront of healthcare in the United States.[1] A recent National Academies of Science report concluded that diagnostic error is responsible for approximately 10% of patient deaths and hospital adverse events.[2] Although diagnostic reasoning—and more broadly, clinical reasoning—is essential to patient care, our understanding of what influences it is limited, making it difficult to mitigate the effects of reasoning errors on patient care.

Clinical reasoning can be defined as the integration of clinical information, medical knowledge, and contextual factors to make patient care decisions.[3–6] A challenging phenomenon in clinical reasoning undoubtedly leading to diagnostic error is *context specificity*: when a physician arrives at two different diagnoses for two different patients who actually have the same symptoms, findings, and, ultimately, the same diagnosis.[7,8] In a recent study of think-aloud reflections on reasoning, for instance, physicians note the presence of contextual factors (e.g., patient affect, diagnostic suggestion), which for some physicians seem to create uncertainty and difficulty with closure of the encounter.[9] We sought to build on this work by using a definition of context grounded in educational theory that could be applied to the rich complexity of practice settings in medicine. Doing so enabled us to empirically explore the phenomenon, one that is a source of unwanted variation in patient care. Thus, we defined case context as—going beyond case *content*—the individual, physical, and social aspects of a patient encounter to include the participants, the setting, and their interactions.[10] In this view, "context is not a fixed set of

surrounding conditions, but a wider dynamical process of which the cognition of an individual is only a part" (p. xii).[11]

We approach the effects of context specificity on clinical reasoning using situated cognition theory, which argues that thinking (here, clinical reasoning) is inextricably bound within the context where it happens (complex interactions among patient, physician, and setting evolving over time).[10,12] Situated cognition recognises the importance of the participants, environment and interactions therein, as noted in the above definition, offering a useful framework for understanding the effects of context specificity. Using this framework, we group contextual factors into those associated with the physician (e.g., fatigue), the patient (e.g., circuitous history), and the clinical environment (e.g., pressure to multitask[9,13]; see Figure 1).

[Insert Figure 1 here]

Another important concept is case specificity, which argues that different content, or different cases (e.g., diagnoses), can lead to different clinical reasoning performance. In explorations of context specificity, something more than case content (i.e., *case specificity*[14]) is influencing clinical reasoning: namely, what we refer to as contextual factors. While contextual factors can *positively* influence clinical reasoning, our work to date and the current study focus on understanding diagnostic error and factors that can *inhibit* it. Recent work suggests contextual factors can inhibit clinical reasoning in both novice and more experienced physicians.[7,15] This aligns with work in psychology suggesting that expert performance is not a stable *trait*, but a shifting, situation-based *state*.[16–18]

Prior work suggests that one mechanism through which contextual factors may affect physicians is mental effort or *cognitive load*: constraints on how many information units one's

working memory can hold and process at a time.[7,19,20] The assumption is that as a clinical encounter becomes more complex, with the introduction of different contextual factors, the associated cognitive load increases, potentially impairing clinical reasoning performance.[7,20] Currently, however, relatively little is known about the relationship between cognitive load and clinical reasoning performance. Combining situated cognition's focus on the interactive elements of the clinical environment with cognitive load theory's focus on individual cognitive management of those elements offers an opportunity to understand how different inhibiting contextual factors may influence clinical reasoning.

Context specificity is recognised as an important problem in medical education, and has been examined through the lens of prototype theory.[21–23] These earlier studies isolated particular elements of the case presentation like the language used, the timing of the presentation of a tentative differential diagnosis, and the familiarity of patient characteristics (e.g., name, age) and examined how they affected participants' (a) determination of a diagnosis and (b) identification of features of a case.[22–30] While this theoretical approach allowed them to tease apart how particular details of case presentations can affect the choice of leading diagnosis, it did not provide a framework for understanding the overall clinical situation in which this diagnostic choice takes place. Moreover, this work and recent work on contextual factors[13,31] have looked only at the endpoint of diagnosis or feature identification, not the broader clinical reasoning *process* (e.g., evidence a physician offers for a given diagnosis) in which these decisions are grounded. Other recent work has taken up a socially situated theoretical model to better understand context specificity, but it has been exploratory and, as such, was conducted with no control group and relatively few participants.[7,13,31–34]

Decrements in clinical reasoning can lead to unwanted variance in performance, patient morbidity and mortality, and/or excessive cost,[2] so we must find innovative ways to examine and

enhance clinical reasoning and context specificity more closely. Thus, the purpose of this study was to investigate both the presence of and the mechanisms behind context specificity, using contextual factors found to be important in our prior work.[7,9,33] Because clinical reasoning performance is related to years of professional experience[3] (our proxy for expertise), we also investigated whether experience affected performance. We also controlled for potential ordering effects. We asked:

(1) Is there a difference in clinical reasoning performance (as measured by open-ended diagnostic questions) when physicians diagnose cases with and without contextual factors?

(2) Is there a difference in self-rated cognitive load (i.e., mental effort) when physicians diagnose cases with and without contextual factors?

We hypothesised that participants would perform better and rate their mental effort as lower in diagnosing cases without inhibiting contextual factors. Further, based on the notion that expertise is a situation-based state (not an invariant trait), we hypothesised that contextual factors would affect physicians equally across experience levels. Finally, because participants regularly see far more than two cases daily, we hypothesised that there would be no ordering effects.

Materials and Methods

To explore how context specificity may impair clinical reasoning, we designed two video simulation cases depicting patients with typical presentations of common diseases: new-onset diabetes mellitus and unstable angina. We believe that videos represent the optimal way to conduct this investigation as videos (widely used training tools) ensure all participants receive an identical “stimulus” to fully control both case content (identical content provided) and potentially relevant contextual factors (i.e., to empirically explore what may underpin context specificity).

They lasted from just under four to six and a half minutes and portrayed a clinical interview, a brief physical exam, and still screens of laboratory findings. We consulted with a group of internal medicine physicians to choose commonly encountered cases and contextual factors that are a part of everyday practice in internal medicine (and that tend to emerge with the types of cases we chose) to mitigate the effects of case (i.e., content) specificity[14] and enhance ecological validity. Prior to filming, the cases were screened by a panel of six medical education experts for authenticity and appropriateness (e.g., typicality).

Study participants were quasi-randomly assigned (based on their study day schedules) to one of two conditions: (a) diabetes case with inhibiting contextual factors (low English proficiency and a patient questioning the physician's credentials); angina case without contextual factors, or (b) angina case with inhibiting contextual factors (misleading diagnostic suggestion, patient reports history circuitously); diabetes case without contextual factors (see Figure 2). Furthermore, because this is early work in this area and because multiple contextual factors are typical across a busy day in practice, we used several contextual factors and two cases. The case content was controlled (i.e., identical presenting symptoms, language and gestures to represent those symptoms, and physical findings) so that it was the same for both diabetes and both angina cases; the only differences were the contextual factors. Also, conditions were balanced to control for potential ordering effects (i.e., whether the contextual factor case came first or second). After watching each case video, participants answered questions about diagnosis and mental effort.

[Insert Figure 2 here]

Participants

A convenience sample (due to the study time demands and institutional requirements of volunteers only) of 39 resident and attending physicians in internal medicine was recruited from the Uniformed Services University of the Health Sciences, Walter Reed National Military Medical Center, and Naval Medical Center San Diego (see Table 1). We sought and received approval from all Institutional Review Boards (complying with World Medical Association Declaration of Helsinki). Participants were allowed to take notes while watching the video cases.

[Insert Table 1 here]

Instruments

Post-Encounter Form

After viewing each video, participants completed a post-encounter form (PEF) asking for: 1) additional information they would like to obtain by history, 2) additional physical exam actions they would take, 3) a problem list, 4) a differential diagnosis, 5) a leading diagnosis, 6) supporting evidence for the diagnosis, and 7) management plans (not discussed in the current manuscript, which focuses on diagnostic reasoning; see Appendix 1 for survey). Participants were given up to 30 minutes (determined to be ample time in prior trials) for completing the items. Items were scored as in prior research where reliability and validity evidence for this instrument were established:[32,33] each free-text response (most participants gave multiple responses for each question) was scored as correct (2 points), partially correct (1 point), or incorrect (0 points) based on a pre-determined scoring key developed by a panel of board certified internists and reported on in prior research[7,32,33] (with reliability between kappa = .82

and kappa = .93 in measure development). Participants were only able to give a single response for the leading diagnosis, but gave multiple responses for the other items. Three of the authors came to complete consensus on the scoring of all new utterances not on the key (less than 3% of responses). Then, in order to compare participants (who gave differing numbers of responses), a percentage correct score was calculated for each of the six items in this study, dividing total number of points received by total number of possible points (e.g., someone who gave five possible differential diagnoses has a total possible raw score of 10 for the differential diagnosis item).

Cognitive Load

Cognitive load was measured through a single self-report item asking participants to: "Select your invested mental effort as you worked through the post-encounter form."[35] The item, used in previous research,[36–38] used a 10-point scale ranging from 1 (very low mental effort) to 10 (very high mental effort).

Data Analysis

To determine if there were differences in clinical reasoning performance, two multivariate analyses of covariance (MANCOVAs) were performed with years of experience (i.e., number of years since medical school graduation) and case order (i.e., whether the diabetes or angina case was first) as covariates. The first between-subjects MANCOVA compared the six PEF scores on the diabetes case with and without contextual factors and the second compared those on the angina case with and without contextual factors. A power analysis indicated that a total sample size of 43 was needed to detect a large-sized effect at a significance level of .05, so the study was slightly underpowered.

To determine if there was a difference in self-reported cognitive load, two ANCOVAs were performed with experience and case order as the same covariates. The first between-subjects ANCOVA compared cognitive load scores on the angina case with and without contextual factors and the second compared cognitive load scores on the diabetes case with and without contextual factors.

Finally, to examine overall score differences, we averaged all PEF items for each case and conducted one-way ANCOVAs.

Results

Across all angina cases, overall mean percentage scores on all PEF items ranged from 43% to 94% (m = 70%). The MANCOVA results revealed significant differences between the conditions with and without contextual factors (Pillai's Trace = .72, F = 12.4, df = [6, 29], $p < .001$, η_p^2 = .72), with no effects for case order or years of experience. Levene's test indicated equal variances for all dependent variables except the supporting evidence item (F = 9.6, p = .006). Follow-up univariate analyses indicated that participants performed significantly worse on all PEF items except additional history questions in the presence of a contextual factor, with large effect sizes[39] (see Table 2, showing the univariate effects). The ANCOVA of overall mean angina scores revealed significant differences between the conditions with and without contextual factors (F = 82.7, df = [1, 34], $p < .001$, η^2 = .71), with a mean score of 77% without contextual factors (SD = .07) and 55% with contextual factors (SD = .1).

[Insert Table 2 here]

Across all diabetes cases, mean percentage scores on PEF items ranged from 37% to 84% (*m* = 70%). MANCOVA results revealed no significant differences between the conditions with and without contextual factors (Pillai's Trace = .33, $F = 2.3$, $df = [6, 28]$, $p = .07$, $\eta_p^2 = .33$) or for either of the covariates. (See Table 3). The ANCOVA of overall mean diabetes PEF score showed no significant differences between conditions ($F = .1$, $df = [1, 33]$, $p = .76$, $\eta^2 = .003$).

[Insert Table 3 here]

There was no significant difference in self-rated cognitive load with or without contextual factors for the diabetes (F [1, 40] = 6.1, p =.38) or angina (F [1, 40] = 1.2, p = .52) cases. Levene's test indicated equal variances for both diabetes (F = .61, p = .61) and angina (F = .55, p = .65). We did, however, observe a trend in cognitive load: self-reported cognitive load was higher in the presence of contextual factors for both diabetes (*m* = 4.85 for no contextual factors and *m* = 5.72 for contextual factors) and angina (*m* = 6.64 for no contextual factors and *m* = 6.95 for contextual factors).

Discussion

In this investigation we sought to experimentally test the phenomenon of context specificity in a group of experienced internists using a theoretically grounded approach. While context specificity in clinical reasoning has become an important area of study,[7–9,25,28,30,33,40–42] to our knowledge, this is the first study to use a robust socially situated theoretical framework, carefully controlled stimulus, ecologically valid measure of clinical reasoning, and fairly large sample of participants. Using typical presentations of common diagnoses, and common contextual factors, we provide empirical evidence for the theoretically predicted negative effects

of context specificity (for the angina case with an aggregated PEF score and with five of six individual PEF measures), with large effect sizes. While prior context specificity work demonstrated significant correlations between PEF items and cognitive load, those studies did not include control groups *without* contextual factors.[9,33] This study extends that work, demonstrating significant performance differences in an angina case across five domains: additional exam actions, problem list, differential diagnosis, leading diagnosis, and supporting evidence. Moreover, this effect held *across* years of experience, indicating the importance of deliberate practice[16,18] and context in the physicians' continuing education: understanding case content is not adequate—physicians must carefully practice reasoning with that content across environments.

Although we found context specificity for the angina case, we did not find such an effect for the diabetes case, despite similar score ranges across the cases and our experts' judgment that both cases were equally common and typical. This finding could be due to inadequate power or to the dose or quality of the contextual factors in the two cases (chosen for their ecological validity with respect to the content area); i.e., perhaps circuitous history poses more or different challenges to clinical reasoning than the agitated non-native English speaker. This could also be evidence of case specificity: some aspect of the content of the diabetes case could make it easier to circumvent contextual effects when compared to the angina case. Thus, neither case content nor contextual factor is the *sole* predictor of clinical reasoning performance. As others have argued, there is no single cause for diagnostic errors, but a nuanced and complex system of interacting conditions.[14,43]

Regarding self-reported cognitive load, while the scores trended in the expected direction for both cases (higher cognitive load with contextual factors), these trends were not statistically significant. This could be due to inadequate power or to problems with our single-item measure

of cognitive load (i.e., inadequate sensitivity). This could also indicate that the cause of the performance decline in the angina case is not the result of increased cognitive load, but due to some other set of factors, such as different emotional reactions.[44] Alternatively, it may be due, in part, to a lack of *awareness* of increased mental effort. For instance, professional gamblers playing with a loaded deck responded physiologically (galvanic skin response) prior to conscious awareness.[45] Thus, individuals may not initially be aware when their cognitive capacity has been exceeded (as our moderate cognitive load scores alongside decreased angina performance would suggest), an area deserving further study. Moreover, since this study focuses explicitly on inhibiting or "disruptive" contextual factors, investigation of factors potentially reducing cognitive load and *improving* performance seems warranted. Further, future work should explore beyond diagnostic reasoning to management reasoning as well.

As the expertise literature would predict, controlling for years of experience did not eliminate the effect of contextual factors.[17,18] In other words, context specificity effects are not limited to newer physicians. Thus, future support tools should be developed not just for residents, but for attending physicians as well. We did note, however, that more experienced physicians performed significantly better on the additional history and additional exam items. Perhaps these diagnostic tasks become more automated or scripted over the years than others (e.g., leading diagnoses). Again, this resonates with prior work indicating that tasks *within* clinical reasoning are equally as important as the broader content or context.[14,43,46]

There are several important study limitations. First, the sample size of 39 participants is only 31% women and, moreover, is relatively small for the statistical test. Yet, this ratio is representative for the participants' institutions and this is actually a large number of participants for a clinical reasoning study (one requiring two hours of physician participant time). Second, we only ran two sets of cases, using different contextual factors in each (diabetes and angina),

potentially affecting the interpretation (i.e., effects could be the result of a *specific type* of contextual factor). Yet, based on our prior work, we wanted to include various contextual factors educators and researchers have hypothesised to be important.

This study has several important practical implications. First, and perhaps most centrally for training and continuing education, these results indicate the importance of identifying areas *other* than content or medical knowledge that contribute to establishing diagnoses as the findings were demonstrated in both residents and attendings. Explicit education in cases with identified contextual factors—perhaps through simulation—could potentially mitigate the negative effects of context specificity. Future work might explore different contextual factors (e.g., appointment length) and their mitigation or elimination. Additionally, the theoretical model and proposed contextual factors could be explored in more authentic settings such as the clinics or wards to better understand context specificity.

Second, this study suggests that we may be underestimating the effect of case specificity on error. Diabetes and angina are two common content areas. While case specificity may explain less variance than individual items within cases in some contexts,[14] it appears to have a significant effect on those items when certain contextual factors are introduced.

Third, while changes in cognitive load were not statistically significant, the trend was in the expected direction, suggesting physicians may be unaware of increased mental effort leading to poorer performance (a finding consistent with our theoretical predictions and prior work[20]). Indeed, if found in larger studies, this finding would suggest that we may not be able to improve diagnostic performance and reduce error without offering strategies for determining when physicians are first "getting out of their depth." Such a conclusion is consistent with the errors literature suggesting that physicians "slow down their thinking when they should," but providing no direction on when or how to do so.[47] Our explorations into context specificity using situated

cognition provide potential insight into why this problem persists despite improvements in awareness of the problem of medical errors, technologic improvements such as electronic health records, and point of care resources for physicians.[48]

Fourth, while serving as an optimal platform for exploring context specificity, the videos themselves may have induced additional cognitive load. We should not assume that clinical reasoning "in person" is the same as clinical reasoning mediated through technology. As such, our findings may have implications for clinical reasoning in technology-enhanced contexts, such as telehealth.

In conclusion, our findings are consistent with expectations of situated cognition theory. Diagnostic error plagues our healthcare system, and we believe that work like this can help illuminate the vexing phenomenon of context specificity. Additionally, this work points to the need for interventions to reduce unwanted performance variance. Such interventions could benefit healthcare systems nationwide.

Author Contributions: All authors collaborated together on the research design and data collection. Konopasky ran the analyses and wrote up the results section. Konopasky and the remaining authors co-wrote the remainder of the paper. All authors offered substantive revisions and approved of the final version of the paper.

Disclaimer: The views expressed in this paper are those of the authors and do not necessarily reflect the official position or policy of the US Government, Department of Defense, Department of the Navy, or the Uniformed Services University.

Research Funding: This study was supported by a grant from the JPC - 1, CDMRP - Congressionally Directed Medical Research Program (#NH83382416)

Competing Interest: The funding organization played no role in the study design; in the collection, analysis, and interpretation of data; in the writing of the report; or in the decision to submit the report for publication.

REFERENCES

1. Singh H, Graber ML. Improving Diagnosis in Health Care--The Next Imperative for Patient Safety. *N Engl J Med*. 2015;373(26):2493.
2. National Academies of Sciences and Medicine E. *Improving Diagnosis in Health Care*. National Academies Press; 2016.
3. Croskerry P. A universal model of diagnostic reasoning. *Acad Med*. 2009;84(8):1022-1028.
4. Elstein AS, Shulman LS, Sprafka SA. Medical problem solving an analysis of clinical reasoning. 1978.
5. Heneghan C, Glasziou P, Thompson M, et al. Diagnostic strategies used in primary care. *Bmj*. 2009;338(apr20_1):b946-b946.
6. Young M, Thomas A, Lubarsky S, et al. Drawing boundaries: the difficulty in defining clinical reasoning. *Acad Med*. 2018;93(7):990-995.
7. Durning SJ, Artino AR, Boulet JR, Dorrance K, van der Vleuten C, Schuwirth L. The impact of selected contextual factors on experts' clinical reasoning performance (does context impact clinical reasoning performance in experts?). *Adv Heal Sci Educ*. 2012;17(1):65-79.
8. Eva KW. What every teacher needs to know about clinical reasoning. *Med Educ*. 2005;39(1):98-106.
9. McBee E, Ratcliffe T, Picho K, et al. Consequences of contextual factors on clinical reasoning in resident physicians. *Adv Heal Sci Educ*. 2015;20(5):1225-1236.
10. Durning SJ, Artino Jr AR, Pangaro LN, van der Vleuten C, Schuwirth L. Perspective: redefining context in the clinical encounter: implications for research and training in medical education. *Acad Med*. 2010;85(5):894-901.

11. Hutchins E. *Cognition in the Wild*. MIT press; 1995.

12. Brown JS, Collins A, Duguid P. Situated cognition and the culture of learning. *Educ Res*. 1989;18(1):32-42.

13. Westbrook JI, Raban MZ, Walter SR, Douglas H. Task errors by emergency physicians are associated with interruptions, multitasking, fatigue and working memory capacity: a prospective, direct observation study. *BMJ Qual Saf*. 2018:bmjqs-2017.

14. Norman G, Bordage G, Page G, Keane D. How specific is case specificity? *Med Educ*. 2006;40(7):618-623.

15. Durning S, Artino Jr AR, Pangaro L, van der Vleuten CPM, Schuwirth L. Context and clinical reasoning: understanding the perspective of the expert's voice. *Med Educ*. 2011;45(9):927-938.

16. Chow DL, Miller SD, Seidel JA, Kane RT, Thornton JA, Andrews WP. The role of deliberate practice in the development of highly effective psychotherapists. *Psychotherapy*. 2015;52(3):337.

17. Ericsson KA. The influence of experience and deliberate practice on the development of superior expert performance. *Cambridge Handb Expert Expert Perform*. 2006;38:685-705.

18. Ericsson KA, Krampe RT, Tesch-Römer C. The role of deliberate practice in the acquisition of expert performance. *Psychol Rev*. 1993;100(3):363.

19. Paas F, Tuovinen JE, Tabbers H, Van Gerven PWM. Cognitive load measurement as a means to advance cognitive load theory. *Educ Psychol*. 2003;38(1):63-71.

20. Young JQ, Van Merrienboer J, Durning S, Ten Cate O. Cognitive load theory: Implications for medical education: AMEE guide no. 86. *Med Teach*. 2014;36(5):371-384.

21. Norman GR, Brooks LR. The non-analytical basis of clinical reasoning. *Adv Heal Sci*

Educ. 1997;2(2):173-184.

22. Hatala R, Norman GR, Brooks LR. Influence of a single example on subsequent electrocardiogram interpretation. *Teach Learn Med*. 1999;11(2):110-117.
23. Kulatunga-Moruzi C, Brooks LR, Norman GR. Using comprehensive feature lists to bias medical diagnosis. *J Exp Psychol Learn Mem Cogn*. 2004;30(3):563.
24. Hatala R, Norman GR, Brooks LR. Impact of a clinical scenario on accuracy of electrocardiogram interpretation. *J Gen Intern Med*. 1999;14(2):126-129.
25. Brooks LR, LeBlanc VR, Norman GR. On the difficulty of noticing obvious features in patient appearance. *Psychol Sci*. 2000;11(2):112-117.
26. Norman GR, Brooks LR, Colle CL, Hatala RM. The benefit of diagnostic hypotheses in clinical reasoning: experimental study of an instructional intervention for forward and backward reasoning. *Cogn Instr*. 1999;17(4):433-448.
27. LeBlanc VR, Norman GR, Brooks LR. Effect of a diagnostic suggestion on diagnostic accuracy and identification of clinical features. *Acad Med*. 2001;76(10):S18-S20.
28. LeBlanc VR, Brooks LR, Norman GR. Believing is seeing: the influence of a diagnostic hypothesis on the interpretation of clinical features. *Acad Med*. 2002;77(10):S67-S69.
29. Young M, Brooks L, Norman G. Found in translation: the impact of familiar symptom descriptions on diagnosis in novices. *Med Educ*. 2007;41(12):1146-1151.
30. Young ME, Brooks LR, Norman GR. The influence of familiar non-diagnostic information on the diagnostic decisions of novices. *Med Educ*. 2011;45(4):407-414.
31. Mercuri M, Sherbino J, Sedran RJ, Frank JR, Gafni A, Norman G. When guidelines don't guide: the effect of patient context on management decisions based on clinical practice guidelines. *Acad Med*. 2015;90(2):191-196.
32. Durning SJ, Artino A, Boulet J, et al. The feasibility, reliability, and validity of a post-

encounter form for evaluating clinical reasoning. *Med Teach*. 2012;34(1):30-37. doi:10.3109/0142159X.2011.590557

33. McBee E, Ratcliffe T, Picho K, et al. Contextual factors and clinical reasoning: differences in diagnostic and therapeutic reasoning in board certified versus resident physicians. *BMC Med Educ*. 2017;17(1):211.

34. ten Cate O, Durning SJ. Understanding Clinical Reasoning from Multiple Perspectives: A Conceptual and Theoretical Overview. 2018:35-46. doi:10.1007/978-3-319-64828-6_3

35. Brunken R, Seufert T, Paas F. Measuring cognitive load. In: Paas J, Moreno R, Brunken R, eds. *Cognitive Load Theory*. London: Cambridge University Press; 2010:181-202.

36. Szulewski A, Gegenfurtner A, Howes DW, Sivilotti MLA, van Merriënboer JJG. Measuring physician cognitive load: Validity evidence for a physiologic and a psychometric tool. *Adv Heal Sci Educ*. 2017;22(4):951-968.

37. Ratcliffe TA, McBee E, Schuwirth L, et al. Exploring Implications of Context Specificity and Cognitive Load in Residents. *MedEdPublish*. 2017;6.

38. Young JQ, Irby DM, Barilla-LaBarca M-L, ten Cate O, O'Sullivan PS. Measuring cognitive load: mixed results from a handover simulation for medical students. *Perspect Med Educ*. 2016;5(1):24-32.

39. Richardson JTE. Eta squared and partial eta squared as measures of effect size in educational research. *Educ Res Rev*. 2011;6(2):135-147.

40. Bowen JL. Educational strategies to promote clinical diagnostic reasoning. *N Engl J Med*. 2006;355(21):2217-2225.

41. Cook DA, Sherbino J, Durning SJ. Management Reasoning: Beyond the Diagnosis. *JAMA*. 2018;319(22):2267-2268.

42. Artino Jr AR, Cleary TJ, Dong T, Hemmer PA, Durning SJ. Exploring clinical reasoning

in novices: a self-regulated learning microanalytic assessment approach. *Med Educ*. 2014;48(3):280-291.

43. Wimmers PF, Fung C. The impact of case specificity and generalisable skills on clinical performance: a correlated traits− correlated methods approach. *Med Educ*. 2008;42(6):580-588.

44. Croskerry P, Abbass A, Wu AW. Emotional influences in patient safety. *J Patient Saf*. 2010;6(4):199-205.

45. Bechara A, Damasio H, Tranel D, Damasio AR. Deciding advantageously before knowing the advantageous strategy. *Science (80-)*. 1997;275(5304):1293-1295.

46. Juma S, Goldszmidt M. What physicians reason about during admission case review. *Adv Heal Sci Educ*. 2017;22(3):691-711.

47. Carol-anne EM, Regehr G, Mylopoulos M, MacRae HM. Slowing down when you should: a new model of expert judgment. *Acad Med*. 2007;82(10):S109-S116.

48. Campione JR, Mardon RE, McDonald KM. Patient Safety Culture, Health Information Technology Implementation, and Medical Office Problems That Could Lead to Diagnostic Error. *J Patient Saf*. 2018.

QUERY FORM

SREP	
Manuscript ID	[Art. Id: 50280]

Journal: SREP

Author:- The following queries have arisen during the editing of your manuscript. Please answer queries by making the requisite corrections at the appropriate positions in the text.

Query No.	Description	Author's Response
Q1:	Please check your article carefully, coordinate with any co-authors and enter all final edits clearly in the eproof, remembering to save frequently. Once corrections are submitted, we cannot routinely make further changes to the article.	
Q2:	Note that the eproof should be amended in only one browser window at any one time; otherwise changes will be overwritten.	
Q3:	Author surnames have been highlighted. Please check these carefully and adjust if the first name or surname is marked up incorrectly. Note that changes here will affect indexing of your article in public repositories such as PubMed. Also, carefully check the spelling and numbering of all author names and affiliations, and the corresponding email address(es).	
Q4:	Please confirm or correct the city/country name inserted in all affiliations.	
Q5:	Please note we have moved the section "Disclaimer" to the end of the methods, as per house style.	

SCIENTIFIC REPORTS
nature research

OPEN

Heart Rate and Heart Rate Variability Correlate with Clinical Reasoning Performance and Self-Reported Measures of Cognitive Load

Received: 29 January 2019
Accepted: 5 September 2019
Published: xx xx xxxx

Soroosh Solhjoo[1], Mark C. Haigney[2], Elexis McBee[3], Jeroen J. G. van Merrienboer[4], Lambert Schuwirth[5], Anthony R. Artino Jr.[6], Alexis Battista[7], Temple A. Ratcliffe[8], Howard D. Lee[9] & Steven J. Durning[10]

Cognitive load is a key mediator of cognitive processing that may impact clinical reasoning performance. The purpose of this study was to gather biologic validity evidence for correlates of different types of self-reported cognitive load, and to explore the association of self-reported cognitive load and physiologic measures with clinical reasoning performance. We hypothesized that increased cognitive load would manifest evidence of elevated sympathetic tone and would be associated with lower clinical reasoning performance scores. Fifteen medical students wore Holter monitors and watched three videos depicting medical encounters before completing a post-encounter form and standard measures of cognitive load. Correlation analysis was used to investigate the relationship between cardiac measures (mean heart rate, heart rate variability and QT interval variability) and self-reported measures of cognitive load, and their association with clinical reasoning performance scores. Despite the low number of participants, strong positive correlations were found between measures of intrinsic cognitive load and heart rate variability. Performance was negatively correlated with mean heart rate, as well as single-item cognitive load measures. Our data signify a possible role for using physiologic monitoring for identifying individuals experiencing high cognitive load and those at risk for performing poorly during clinical reasoning tasks.

Diagnostic accuracy and the precise development of a management plan are imperative to improving patient safety[1–3]. Clinical reasoning can be defined as the cognitive steps (e.g. information gathering, problem representation, generating and refining diagnostic hypotheses) leading up to and arriving at a diagnosis and a management plan[4–6]. Assessing the clinical reasoning performance, however, is challenging due to the limitations of the assessment methods, many of which do not incorporate the complexity and contextual nature of clinical reasoning as a construct[7]. Given the notion that clinical reasoning is at the heart of what it means to be a clinician[8], it is essential that we enhance our understanding of clinical reasoning and how it can be assessed.

Q1 Q2 Q3 Q4

Cognitive load theory can be a useful explanatory theoretical lens for better understanding of when clinical reasoning is successful and when it goes wrong. Cognitive load theory posits that working memory is limited in both capacity and duration (i.e., only a few elements of information can be processed at any given time, and

[1]Division of Cardiovascular Pathology, Johns Hopkins University School of Medicine, Baltimore, USA. [2]Professor of Medicine and Pharmacology, Director of Cardiology, USUHS, Bethesda, USA. [3]Associate Professor of Medicine, USUHS, Bethesda, USA. [4]Professor of Medical Education, Maastricht University, Maastricht, The Netherlands. [5]Strategic Professor in Medical Education, Flinders University, Adelaide, Australia. [6]Professor of Medicine, Deputy Director, Division of Health Professions Education, USUHS, Bethesda, USA. [7]Assistant Professor of Medicine, USUHS, Bethesda, USA. [8]Associate Professor of Medicine, UT Health San Antonio, San Antonio, USA. [9]San Antonio Uniformed Services Health Education Consortium, Bethesda, USA. [10]Professor of Medicine, Director, Division of Health Professions Education, USUHS, Bethesda, USA. Correspondence and requests for materials should be addressed to S.S. (email: soroosh@jhmi.edu)

under realistic circumstances, are held in working memory for less than twenty seconds)[9,10]. Cognitive load refers to one's mental effort to complete a task, and it is primarily driven by element interactivity; that is, the number of cognitive elements that are simultaneously processed in working memory. In education studies, cognitive load theory posits three types of cognitive load which are affected differently by instruction and consequently have different implications for learning and performance: intrinsic load, determined by the task's complexity and the learner's prior knowledge; germane load, the cognitive load of construction and automation of schemata in long-term memory; and extraneous load, generated by the manner in which information is presented to learners interfering with schema acquisition and automation[11–15].

Some researchers have criticized[16,17] cognitive load theory as being difficult to falsify experimentally. Indeed, several studies tested hypotheses concerning the overall cognitive load and then interpreted their results in terms of intrinsic, extraneous and germane cognitive loads. This is problematic; hypotheses should be prespecified in terms of the different types of cognitive load[15,18]. There is no a priori reason that explanatory theoretical constructs cannot be applied to create a deeper understanding of complex phenomena and provide a foundation for the formulation of more concrete hypotheses[19,20].

Assessing the Impact of Cognitive Load on Clinical Reasoning

Several studies have provided evidence on the reliability and the validity of self-reported measures of cognitive load[15,18]. Clinical reasoning performance may be negatively associated with high cognitive load[21]. Further, high cognitive load may contribute to context specificity; e.g., seeing two patients with the same chief complaint, symptoms and findings and yet coming to different diagnoses[22]. Nonetheless, self-reported measures of perceived cognitive load may provide an incomplete picture of cognitive load[23]. Part of the problem is that individuals may be unaware of when their cognitive load exceeds capacity[21]. This is particularly salient when the excessive cognitive load happens in the "here and now" – during the busy daily clinical practice. It is reasonable to assume that when excessive cognitive load occurs, the clinician does not have cognitive resources left to reflect on the balance between cognitive load and capacity. Thus, especially in those situations, self-reports are logically of limited practical value.

By contrast, physiologic measurements are less likely to be influenced by the limitations of an individual's ability to self-assess cognitive load. In particular, cardiovascular measures may be suitable indices of cognitive load due to their reliability and the feasibility of continuous recording[24]. One well-established cardiovascular measure is heart rate variability. Changes in heart rate variability indicate modulation of the autonomic nervous system mainly in response to changes in blood pressure and mental stress[25]. Importantly, the performance of subjects under stress may be positively or negatively impacted by the autonomic nervous system, and so, poorly controlled autonomic tone may contribute to poor performance and be a target for intervention. The connection between heart rate variability and cognitive function has been the subject of several studies[26–30]. Although heart rate variability is commonly used as an index of autonomic nervous system activity, it remains unclear whether it is sufficiently sensitive to variations in cognitive load in education scenarios. For example, in an exploratory study of computer-based training strategies, Paas *et al*.[31] found no correlation between cognitive load and one specific aspect of heart rate variability; however, they only looked at the spectral power in the low frequency band (i.e., spectral power of the frequency band of 0.07–0.14 Hz), did not parse out different components of cognitive load, and did not include an orthogonal measure of the impact of cognitive load on the autonomic nervous system. For further review of the studies investigating the use of physiologic measures, particularly heart rate variability, to assess stress and mental workload, please see refs[32–34].

Furthermore, research conducted in other domains suggests that biological changes may precede cognitive awareness when individuals are struggling with their thought processes (i.e., high cognitive load); for example, among professional gamblers, high sympathetic tone, as measured by skin galvanic response, was observed before these professionals could vocalize a problem with a fixed card deck[35]. We therefore specifically sought to explore if this phenomenon is present in the context of clinical reasoning as improving physician's awareness of when help is needed could dramatically improve care and reduce error.

The purpose of this exploratory study is first to determine whether cardiovascular measures can be used as markers for cognitive load and, second, to investigate whether the more feasible option of self-report measures have biological validity evidence for clinical reasoning performance in medical students. Here, in addition to measuring the spectral power in different frequency bands, we use time-domain measures of heart rate variability; i.e., the root mean square of differences of successive heartbeat intervals (RMSSD), and the standard deviation of the normal to normal heart beat intervals (SDNN), which assesses total variability and makes no prior assumptions about the specific frequency band likely to be affected[25]. Moreover, we also measure the total variability of the QT interval (the period between the beginning of the Q wave and the end of the T wave in each cycle of the ECG signal) as an orthogonal index of the impact of cognitive load on the autonomic nervous system. This measure of QT variability is an index of the effects of changes in autonomic tone on the heart rhythm. Because they are objective and reliable[24], physiologic markers could potentially provide an effective means to investigate the validity of self-reported measures of cognitive load.

We predicted that our findings would not only detect an association between cognitive load and clinical reasoning performance consistent with our theoretical framework, but also that there would be an association between cognitive load measures and sympathetic tone, providing additional evidence for the validity of cognitive load self-reported measures. We further predicted that these associations would be detectable during three episodes of relatively mundane clinical reasoning and not be restricted to extraordinarily challenging encounters.

Methods

Participants. Fifteen third- and fourth-year medical students from the Uniformed Services University of the Health Sciences were recruited to view three videos depicting physician-patient interactions and then complete a post-encounter form (PEF) for each one. Their ECG was recorded using a Holter monitor starting 24 hours before (baseline) and while they watched the videos and reported their clinical reasoning (test). Holter data for five of the participants were excluded from analysis for the following reasons: for one participant, the recording was too noisy; for two, the time stamps were not available; and for two, the data were not recorded for the full period of the experiment. The data of the remaining 10 participants were used in the analysis. There were no exclusion criteria.

Assessment of clinical reasoning performance. As a first step, several authors crafted a written script for three video-based cases. The cases were then reviewed by an expert panel of eight internal medicine physicians and modifications were made to the script. Video cases were then filmed and re-reviewed by the same expert panel of eight internal medicine physicians for consistency.

Next, the PEF scoring rubric was constructed based on the script by having the authors generate answers for the different sections of the PEF. This was followed by review of the answers by the entire panel of experts. Following two rounds of reviews, we were able to establish complete consensus for correct, partially correct, and incorrect responses for each section. After having participants complete the PEF, additional answer options were generated that were not a part of the key (note: less than 2% of answers were not on the original key). These answers were reviewed by four of the study authors and complete consensus was reached for final responses. Reliability and validity evidence for use of the PEF has been collected previously[36].

Procedures. After informed consent, a trained researcher fitted participants with a 12-lead Holter recorder 24 hours prior to the test to establish a baseline reading. Following the 24-hour baseline period, participants were asked to sit behind a computer desk and view three outpatient clinical encounter videos that had previously undergone expert review. The first video portrayed a diagnosis of an acute retroviral syndrome, the second patient presented with colorectal cancer and an acute pulmonary embolism, and the third patient presented with new onset diabetes. The second case video, representing a life-threatening presentation, was anticipated to lead to the greatest amount of cognitive load and sympathetic tone due to the acuity of the presentation. We did not include measures of empathy, anxiety, or emotional stress as these cases were typical for the work that these physicians would be expected to encounter in practice.

During the test period, for each video, participants viewed the video and then completed the PEF followed by a single-item cognitive load rating scale. Participants then immediately re-watched the video and were asked to explain their reasoning orally using a think-aloud protocol that is similar to cued retrospective reporting[37]. Following these steps, participants completed a 10-item cognitive load measure one time at the end of the test.

Cognitive load measures. After completion of each PEF, participants provided a self-reported single-item cognitive load measure[31]. For this, they rated their level of cognitive load exerted on the task using a Likert-type scale ranging from 1 (no cognitive load exerted) to 9 (very high cognitive load). This single-item measure is brief and has been used in several prior studies[38].

An additional self-reported measure of cognitive load was given to each participant at the end of the three cases. It consisted of a 10-item questionnaire designed to measure the three different types of cognitive load (extraneous, germane, and intrinsic). We included these measures given the reported limitations of the single-item cognitive load measure[39]. All questionnaire items use an 11-point Likert-type scale that ranged from 0 to 10, with higher scores indicating higher cognitive load. Validity of the scores on this questionnaire as a psychometric measure has been shown in domains outside medical education[15,18].

Physiologic measures. ECG recordings were obtained using a high-resolution (1 kHz), digital, 12-lead, portable Holter monitoring system (Mortara Instrument Inc., Milwaukee, WI) starting 24 hours prior to the test and during the intervention. Several time and frequency domain measures were extracted from each participant's ECG according to established guidelines[25]. Time domain measures consisted of the mean heart rate (HR, beats/min), heart rate variability calculated as the standard deviation of the time between normal beats (SDNN, msec) and root mean square of successive differences of heartbeat intervals (RMSDD, msec). The power of heart rate variability time series was measured in three frequency bands: very low frequency (VLF; 0.0037–0.04 Hz), low frequency (LF; 0.04–0.15 Hz), and high frequency (HF; 0.15–0.4 Hz). LF is associated with combined vagal and sympathetic stimulations[40] and HF is associated with vagal stimulation and the respiratory system's effect on the heart rate[41]; therefore, these two measures are not independent.

On average, each task took 7.01 ± 2.13 min (mean $\pm$ standard deviation), and the shortest task across all participants lasted 4.5 minutes. Therefore, to account for all the tasks in the test, the analysis was performed on 4.5-minute segments of the ECG signal, using a moving window at 0.5-minute steps. For each task, we used the average of the parameters calculated for each of the windows covered during that task. For example, VLF reported for a 7-min task is the average of VLF calculated for each of the six 4.5-min windows covered during that task. This would improve parameter estimates and lower distortion.

The QT interval was measured using a semi-automated, template matching algorithm that has been previously described[42]. Briefly, the algorithm generates several signal-averaged templates from a chosen ECG lead. For each template, the investigator identifies a representative complex, including the entire QT and U wave in order to include all components related to depolarization and repolarization of the ventricles. The inclusion of the U wave has been previously shown to improve the predictive value of the metric for life-threatening arrhythmias[43]. Each individual QT interval value is then calculated as how much each beat needs to be stretched or compressed to

fit the corresponding template QT. A normalized QT variability index (QTVI) was also derived according to the following equation:

$$\mathrm{QTVI} = \log_{10}[(\mathrm{QT}v/\mathrm{QT}^2)/(\mathrm{HR}v/\mathrm{HR}^2)],$$

where HR = mean heart rate, HRv = heart rate variance, QT = mean QT interval, and QTv = QT interval variance. QTVI formula is designed to produce an independent measure by including QT and HR (which are not independent) in the numerator and the denominator.

To limit the effect of posture or physical activity on the physiologic measures, the participants were asked to keep sitting as they watched the videos, filled out the questionnaires, or explained their thinking process.

Clinical reasoning performance measures. Participants' performance for each scenario was measured using a PEF, on which they indicated a leading diagnosis, differential diagnosis, supporting data and a therapeutic management plan. Scoring of the PEF entailed having a group of experts construct and revise answer key responses through a series of discussions. Complete consensus was achieved for this scoring rubric. Reliability and validity of this PEF for the assessment of clinical reasoning has been previously established[36,44]. Each PEF consisted of the following prompts:

Patient history. What else do you want to ask this patient? (List one to five questions).

Physical exam. What else would you want to look for on this patient's physical exam? (List one to five items).

Differential diagnosis. What is your differential diagnosis? (please list in order of likelihood and list at least 3 responses).

Supporting evidence. What data supports this diagnosis? (List one to five pieces of evidence).

Treatment/management plan. What is your treatment/management plan for this patient (diagnostic and/or therapeutic).

An expert panel generated scores for every entry on the PEF with complete consensus. This was achieved after two rounds of review and edits to potential PEF responses. Scores for each response ranged from 0 (incorrect), to 1 (partially correct), and 2 (correct). Scores for all responses were tallied to generate a total score for clinical reasoning performance (maximum score of 30).

Data analysis. Correlation analysis was performed to assess the association between the self-reported cognitive load measures and physiologic measures. For this purpose, partial correlation was measured to control for gender differences in physiologic measures of heart rate variability[33]. Correlation analysis was also used to explore the relationship between clinical reasoning performance scores and cognitive load using both physiologic measures and self-reported measures of cognitive loads. We extracted the time and frequency domain parameters for the time period that each participant spent watching and completing the PEF and think-aloud protocols for each video. Participants' average physiologic measures during each task were used to calculate the correlation coefficients. Signal processing, feature extraction and data analysis were performed using in-house software developed in MATLAB[45]. Data are presented as mean ± standard error of the mean unless noted otherwise. For correlation analysis, we set type I error rate of $\alpha = 0.05$. When considering each task separately, we set the minimum correlation coefficient of $|\rho| \geq 0.67$. With 15 subjects, our analysis would have 80% power (i.e., type II error rate of $\beta = 1 - \text{power} = 0.2$). Because we lost data from 5 out of 15 subjects, our analysis power dropped to 60% (i.e., $\beta = 0.4$)[46]. Due to this increase in type II error, there might be associations between the cardiovascular parameters of each specific task and performance/cognitive load measures that we failed to detect; however, the type I error rate was kept low ($\alpha = 0.05$).

Ethical approval. The data were stored and analyzed anonymously, and this study was deemed exempt by IRB at Uniformed Services University of the Health Sciences. Informed consent was obtained from all participants prior to the study. All research was performed in accordance with relevant guidelines and regulations.

Q5 **Disclaimer.** The views expressed in this paper reflect the opinions of the authors only and not the official policy of the United States Army, Uniformed Services University, or the Department of Defense.

Results

Study cohort. The final sample of ten participants contained 2 females, and the mean age was 25. None were on any prescribed medications. Average electrocardiographic variables recorded at baseline (24-hour period preceding the test) and during the test are reported in Table 1. Participants took 5.6 ± 0.2 min to watch the video, 8.97 ± 0.32 min to fill out the PEF, and 6.5 ± 0.33 min for think-aloud for each case.

Cognitive load and clinical reasoning performance. Participants' performance scores on the PEF ranged from 11 to 25 (17 ± 1.73) for the first video, 16 to 27 (22.1 ± 1.29) for the second video, and 10 to 25 (16.7 ± 1.57) for the third video. The average single-item measures of cognitive load were 5.9 ± 0.53 after the first (CL1), 6.5 ± 0.4 after the second (CL2), and 7.4 ± 0.31 after the third video (CL3), showing a steady increase (CL3 > CL1, $p < 0.05$). On the 10-item inventory, intrinsic, germane and extraneous types of cognitive load were measured: scores for intrinsic and germane cognitive loads ranged from 3 to 8 (4.97 ± 0.55 and 5.03 ± 0.52, respectively), and scores for extraneous cognitive load ranged from 0 to 10 (2.07 ± 0.98).

	HR (beat/min)	SDNN (msec)	QTVI
24 hours prior to the test	71.04 ± 2.73	79.11 ± 5.94	−1.42 ± 0.06
During the test	68.88 ± 2.70	70.32 ± 3.80	−1.46 ± 0.10

Table 1. Baseline mean values of the physiologic parameters measured 24 hours prior to the test. The parameters are reported as mean ± standard error of the mean.

Cognitive Load Measure	Physiologic Measure	Correlation Coefficient	*p*-value
Intrinsic	LF	0.91	0.001
Intrinsic	SDNN	0.71	0.031
Intrinsic	RMSSD	0.69	0.040
Germane	VLF	0.68	0.045
CL1 + CL2 + CL3	QTVI	0.72	0.030
Intrinsic	t1 LF	0.70	0.035
Intrinsic	v2 LF	0.77	0.016
Intrinsic	v2 QT	0.75	0.033
Intrinsic	p2 LF	0.72	0.028
Intrinsic	p2 RMSSD	0.71	0.032
Intrinsic	t2 SDNN	0.76	0.019
Intrinsic	t2 VLF	0.73	0.027
Intrinsic	t2 LF	0.73	0.026
Intrinsic	v3 LF	0.73	0.026
Intrinsic	p3 LF	0.74	0.022
Intrinsic	p3 RMSSD	0.74	0.023
Intrinsic	t3 SDNN	0.90	0.001
Intrinsic	t3 VLF	0.76	0.018
Intrinsic	t3 LF	0.90	0.001
Intrinsic	t3 HF	0.72	0.030
Intrinsic	t3 RMSSD	0.86	0.003
Germane	p1 SDNN	0.84	0.005
Germane	p1 VLF	0.82	0.007
Germane	p1 LF	0.72	0.028
CL2	p2 QTVI	0.81	0.008
CL2	t2 QTVI	0.77	0.016
CL2	p3 QTVI	0.89	0.001
CL2	t3 QTVI	0.81	0.008
CL3	p3 RMSSD	0.69	0.040

Table 2. Correlations between measures of self-reported cognitive load and physiologic measures. v*n*, p*n* and t*n* indicate the physiologic measures averaged during watching, PEF completion and the think-aloud sessions for clinical case *n* (1–3), respectively. When task number is not indicated, the full test period (63.1 ± 1.87 min) was used for the measurement.

Across all three case videos (n = 30), performance scores negatively correlated with single-item measures of cognitive load ($r = -0.47$, $p < 0.01$). However, we did not find any statistically significant correlation between the 10-item measures of the three different types of cognitive load and performance scores.

Cognitive load and cardiovascular measures. Here, we assessed the correlation between measures of cognitive load (intrinsic, germane, and extraneous) and cardiovascular measures. During the test, intrinsic cognitive load was positively correlated with heart rate variability features in both time and frequency domains, including SDNN, RMSSD, LF and VLF power (Table 2). Of note, the correlation between self-reported intrinsic cognitive load and SDNN measured during think-aloud sessions increased across the three video tasks (Fig. 1). A steady increase was also seen in the correlation between self-reported intrinsic cognitive load and LF power measured during think-aloud (Table 2). QTVI was strongly associated with single-item measures of cognitive load during the second case video (Table 2).

Clinical reasoning performance and cardiovascular measures. Table 3 lists the statistically significant correlations of clinical reasoning performance scores with cardiovascular measures. Performance scores for clinical case 2 were positively correlated with heart rate variability as measured by SDNN and VLF while the

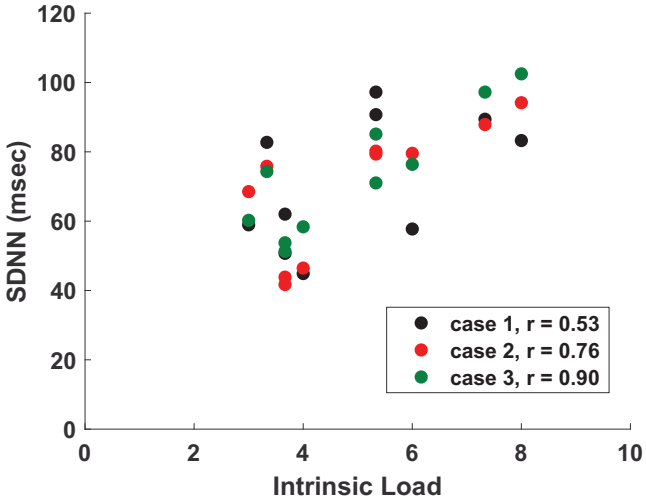


Figure 1. Scatterplot of SDNN versus intrinsic cognitive load during the think-aloud sessions of the three clinical cases. The correlation between intrinsic cognitive load and SDNN increased during the experiment.

PEF-based Performance Measure	Physiologic Measure	Correlation Coefficient	*p*-value
Case 2 Performance Score	v2 SDNN	0.69	0.042
	v2 VLF	0.68	0.042
	v2 QT	0.71	<0.05
	v3 QT	0.68	0.045
	p3 HR	−0.70	0.037
	p3 QT	0.69	0.042
	t3 QT	0.73	0.026
Case 3 Performance Score	v3 QTVI	−0.73	0.024

Table 3. Correlations between objective performance measures and cardiovascular variables. v*n*, p*n* and t*n* indicate the physiologic measures averaged during watching the video, PEF completion and the think-aloud sessions for case *n* (1–3), respectively.

participants watched the video. Those with better performance for case 2, manifested lower heart rates and higher QT interval during the following task. These correlations were not present for the other two cases. For case 3, the performance score was negatively correlated with QTVI while the participants watched the video.

Discussion

The major findings of this study were threefold: (a) we found strong correlations between cardiovascular measures and self-reported measures of cognitive load during clinical reasoning; (b) performance scores negatively correlated with single-item measures of cognitive load; and (c) we found strong negative correlations between objective measures of performance and mean heart rate for one task. QT duration was also correlated with performance, but this effect likely reflects the same phenomenon as heart rate, i.e. an increase in sympathetic tone. The correlations between performance and the physiologic measures reported in this study were not present with the physiologic measurements 24 hours prior and were only present on the test day. These findings were consistent with our hypothesis that high cognitive load would correlate with physiologic measures of sympathetic tone. The correlations were strongest for case 2, which represented the most urgent presentation (a patient with colorectal cancer and a pulmonary embolism), as the acuity of this life-threatening presentation would be expected to invoke greater sympathetic tone. This finding could have significant implications for the assessment of individuals performing complex tasks that are associated with significant failure risks.

Our analysis showed a positive correlation between intrinsic cognitive load and heart rate variability frequency and time domain measurements (Table 2). An increase in heart rate variability is generally regarded as an indication of a robust autonomic nervous and cardiovascular system[25], and it is somewhat surprising that it was associated with increased cognitive load. Heart rate variability can increase due to an increase in parasympathetic or sympathetic tone (or both)[40]. Our findings suggest that an increase in perceived cognitive load appears to result in an increase in both sympathetic and parasympathetic components of the autonomic nervous system. While mental stress is typically associated with decreased parasympathetic tone, an increase in blood pressure may have had the opposite effect in our cohort. Mental stress has been shown to increase blood pressure[47], and an increase in blood pressure in young healthy individuals could be expected to increase parasympathetic tone via the baroreceptor reflex mechanism.

QTVI, a validated measure predicting adverse cardiovascular events, was correlated with self-reported single-item measures of cognitive load overall, and particularly at the end of the second case, as well as the performance scores for the third case. In healthy individuals, heart rate and QT interval are inversely correlated; as heart rate increases, the QT interval shortens. Activation of the sympathetic nervous system and parasympathetic withdrawal significantly increases heart rate and shortens the QT interval through direct and indirect effects on the myocardium. QTVI is a log ratio of normalized QT variance over normalized heart rate variance, and therefore an increase in QTVI in the setting of increased heart rate variability is somewhat unusual, indicating that repolarization variability increased to a greater extent than heart rate variability. Identifying trainees who are experiencing increased cognitive load could have important implications for physician health and for program level wellness initiatives.

Clinician's performance is a critical concern to patients and health systems, and identifying clinicians that are in danger of clinical reasoning performance failure prior to making an error is an important goal. In this study, we found a strong inverse correlation between heart rate and an objective performance score during a clinical reasoning exam, indicative of activation of the sympathetic nervous system in those at risk of doing poorly. In addition, self-reports of cognitive load are not feasible to be used during normal clinical practice, whereas these are the contexts in which this balance between cognitive load and capacity may be most detrimental. If a clinician is overwhelmed at times by the situational demands, they generally do not have the time to sit and think, or reflect, or take a 'timeout' in every situation. The understanding of the relationship between cognitive load, risk of underperformance and physiological parameters may be useful to design monitoring warning instruments for practicing clinicians in complex settings to enhance self-monitoring – a critical component of self-regulation.

The current study is unique in that it bridges multiple fields: cognitive psychology, physiology, and medicine. It is a first attempt to measure clinical reasoning performance using the proxy of cognitive load with physiologic parameters that are not subject to error in self-reports. As stated in the introduction, all assessments bear in them the problem of having to infer mental processes from observing external behavior and this inference is always influenced by the validity evidence in the context of current validity theory. Physiological parameters could potentially serve as a more direct measurement of cognitive load. Therefore, the findings from this study may have important practical significance and implications in medical education, especially with respect to the development of tools to optimize the influence of cognitive load and improve clinical reasoning performance. The increasing use of personalized monitors for heart rate and even electrocardiogram makes it likely that these findings could be potentially employed to monitor trainees to optimize their clinical reasoning ability, as well as their personal health and to preempt clinical failure.

This study also had several limitations. First, the sample of participants in this study was quite small. Out of original 15 participants, five participants' data had to be excluded for technical reasons. However, the identified effect sizes were large, and the results were statistically significant. Second, the study was conducted in a low-stakes experimental environment, which might have attenuated the effects of cognitive load on performance. Third, the absence of blood pressure as a gauge of physiologic response to stress limits any inferences we might have been able to make regarding its potential moderating role on the impact of cognitive load on performance. Fourth, we did not explore the learning process in this investigation, and there may be differential effects on learning and performance in trainees in terms of cognitive load.

For the purposes of our analysis, we have applied the prevalent assumption that the autonomic nervous system – and the indices of heart rate variability and QTVI – represent purely reactive phenomena triggered by the perception of external stimuli. The "Polyvagal Theory", however, suggests that there are phylogenetic differences in the organization of the parasympathetic system that support a bidirectional interaction for the autonomic system and higher behaviors[48]. In mammals, the parasympathetic system incorporates central nuclei that allow the system to not only suppress sympathetically-driven vegetative functions (i.e., blood pressure and heart rate), but to also modulate internal perceptions, facial behaviors, and ultimately social interactions[49]. Testing this hypothesis is beyond the scope of this study, but future investigations could explore the impact of parasympathetic intervention (i.e., exercise training) on perceived cognitive load and performance.

Our current findings have the potential to inform assessment of clinical reasoning performance in authentic (e.g. patient care) settings. Such work could also advance our understanding of context specificity, which leads to unwanted variation in physician performance. For example, consistent with the literature on cognitive load, instructional materials could then be developed to assist the clinician/student with reducing cognitive load and improving future performance. The inclusion of physiologic monitoring in a training regime could provide "real time" feedback to the learner regarding the effectiveness of that regime.

One implication for practice is to determine if expected increases in sympathetic tone would be seen before an individual is able to vocalize that they are dealing with a challenging situation (e.g., that they are "out of their depth"). We envision future means of looking at heart rate variability by emerging hand-held or wearable technologies to help the physician know when they may need help with clinical care, as well as using heart rate variability monitors to generate validity evidence for more common assessment measures of clinical reasoning in practice.

Data Availability

Anonymized data are available from the corresponding author upon reasonable request.

References

1. The National Academies of Sciences, Engineering, and Medicine. *Improving diagnosis in health care*. (National Academies Press, 2016).
2. Donaldson, M. S., Corrigan, J. M. & Kohn, L. T. *To err is human: building a safer health system*. Vol. 6 (National Academies Press, 2000).
3. Makary, M. A. & Daniel, M. Medical error—the third leading cause of death in the US. *BMJ* **353**, i2139, https://doi.org/10.1136/bmj.i2139 (2016).

4. Durning, S. J., Artino, A. R. J., Pangaro, L. N., van der Vleuten, C. & Schuwirth, L. Perspective: Redefining Context in the Clinical Encounter: Implications for Research and Training in Medical Education. *Academic Medicine* **85**, 894–901, https://doi.org/10.1097/ACM.0b013e3181d7427c (2010).
5. Eva, K. W. What every teacher needs to know about clinical reasoning. *Medical education* **39**, 98–106 (2005).
6. Daniel, M. *et al*. Clinical Reasoning Assessment Methods: A Scoping Review and Practical Guidance. *Academic Medicine* **94**, 902–912, https://doi.org/10.1097/acm.0000000000002618 (2019).
7. Schuwirth, L. Is assessment of clinical reasoning still the Holy Grail? *Medical education* **43**, 298–300, https://doi.org/10.1111/j.1365-2923.2009.03290.x (2009).
8. Durning, S. J. *et al*. The impact of selected contextual factors on experts' clinical reasoning performance (does context impact clinical reasoning performance in experts?). *Advances in health sciences education* **17**, 65–79 (2012).
9. Peterson, L. & Peterson, M. J. Short-term retention of individual verbal items. *Journal of experimental psychology* **58**, 193 (1959).
10. Goldstein, E. *Cognitive psychology: Connecting mind, research and everyday experience*. (Nelson Education, 2010).
11. Paas, F., Renkl, A. & Sweller, J. Cognitive load theory and instructional design: Recent developments. *Educational psychologist* **38**, 1–4 (2003).
12. Van Merrienboer, J. J. & Sweller, J. Cognitive load theory and complex learning: Recent developments and future directions. *Educational psychology review* **17**, 147–177 (2005).
13. Leppink, J. & van den Heuvel, A. The evolution of cognitive load theory and its application to medical education. *Perspectives on Medical Education* **4**, 119–127, https://doi.org/10.1007/s40037-015-0192-x (2015).
14. Leppink, J., Gog, T., Paas, F. & Sweller, J. Cognitive load theory: researching and planning teaching to maximise learning. *Researching medical education* (eds Jennifer Cleland & Steven J. Durning) 207–218 (John Wiley & Sons, Ltd, 2015).
15. Leppink, J., Paas, F., Van der Vleuten, C. P. M., Van Gog, T. & Van Merriënboer, J. J. G. Development of an instrument for measuring different types of cognitive load. *Behavior Research Methods* **45**, 1058–1072, https://doi.org/10.3758/s13428-013-0334-1 (2013).
16. Moreno, R. When worked examples don't work: Is cognitive load theory at an impasse? *Learning and Instruction* **16**, 170–181 (2006).
17. De Jong, T. Cognitive load theory, educational research, and instructional design: some food for thought. *Instructional science* **38**, 105–134 (2010).
18. Leppink, J., Paas, F., Van Gog, T., van Der Vleuten, C. P. & Van Merrienboer, J. J. Effects of pairs of problems and examples on task performance and different types of cognitive load. *Learning and Instruction* **30**, 32–42 (2014).
19. Cianciolo, A. T., Eva, K. W. & Colliver, J. A. Theory development and application in medical education. *Teaching and learning in medicine* **25**, S75–S80 (2013).
20. Gerjets, P., Scheiter, K. & Cierniak, G. The scientific value of cognitive load theory: A research agenda based on the structuralist view of theories. *Educational Psychology Review* **21**, 43–54 (2009).
21. Durning, S., Artino, A. R., Pangaro, L., van der Vleuten, C. P. & Schuwirth, L. Context and clinical reasoning: understanding the perspective of the expert's voice. *Medical Education* **45**, 927–938 (2011).
22. Eva, K. W. On the generality of specificity. *Medical education* **37**, 587–588 (2003).
23. Choi, H.-H., Van Merriënboer, J. J. & Paas, F. Effects of the physical environment on cognitive load and learning: towards a new model of cognitive load. *Educational Psychology Review* **26**, 225–244 (2014).
24. Croizet, J.-C. *et al*. Stereotype Threat Undermines Intellectual Performance by Triggering a Disruptive Mental Load. *Personality and Social Psychology Bulletin* **30**, 721–731, https://doi.org/10.1177/0146167204263961 (2004).
25. Electrophysiology, T. F. o. t. E. S. o. C. t. N. A. S. o. P. a. E. Heart Rate Variability: Standards of Measurement, Physiological Interpretation, and Clinical Use. *Circulation* **93**, 1043–1065, https://doi.org/10.1161/01.cir.93.5.1043 (1996).
26. Thayer, J. F. & Lane, R. D. Claude Bernard and the heart–brain connection: Further elaboration of a model of neurovisceral integration. *Neuroscience & Biobehavioral Reviews* **33**, 81–88 (2009).
27. Luque-Casado, A., Perales, J. C., Cárdenas, D. & Sanabria, D. Heart rate variability and cognitive processing: The autonomic response to task demands. *Biological psychology* **113**, 83–90 (2016).
28. Spangler, D. P., Gamble, K. R., McGinley, J. J., Thayer, J. F. & Brooks, J. R. Intra-individual variability in vagal control is associated with response inhibition under stress. *Frontiers in human neuroscience* **12**, 475, https://doi.org/10.3389/fnhum.2018.00475 (2018).
29. Castaldo, R., Montesinos, L., Melillo, P., James, C. & Pecchia, L. Ultra-short term HRV features as surrogates of short term HRV: a case study on mental stress detection in real life. *BMC Medical Informatics and Decision Making* **19**, 12, https://doi.org/10.1186/s12911-019-0742-y (2019).
30. Delliaux, S., Delaforge, A., Deharo, J.-C. & Chaumet, G. Mental Workload Alters Heart Rate Variability, Lowering Non-linear Dynamics. *Frontiers in Physiology* **10**, https://doi.org/10.3389/fphys.2019.00565 (2019).
31. Paas, F. G., Van Merriënboer, J. J. & Adam, J. J. Measurement of cognitive load in instructional research. *Perceptual and motor skills* **79**, 419–430 (1994).
32. Castaldo, R. *et al*. Acute mental stress assessment via short term HRV analysis in healthy adults: A systematic review with meta-analysis. *Biomedical Signal Processing and Control* **18**, 370–377 (2015).
33. Kim, H.-G., Cheon, E.-J., Bai, D.-S., Lee, Y. H. & Koo, B.-H. Stress and heart rate variability: A meta-analysis and review of the literature. *Psychiatry investigation* **15**, 235 (2018).
34. Charles, R. L. & Nixon, J. Measuring mental workload using physiological measures: a systematic review. *Applied ergonomics* **74**, 221–232 (2019).
35. Bechara, A., Damasio, H., Tranel, D. & Damasio, A. R. Deciding advantageously before knowing the advantageous strategy. *Science* **275**, 1293–1295 (1997).
36. Durning, S. J. *et al*. The feasibility, reliability, and validity of a post-encounter form for evaluating clinical reasoning. *Medical teacher* **34**, 30–37 (2012).
37. van Gog, T., Paas, F., van Merriënboer, J. J. G. & Witte, P. Uncovering the Problem-Solving Process: Cued Retrospective Reporting Versus Concurrent and Retrospective Reporting. *Journal of Experimental Psychology: Applied* **11**, 237–244, https://doi.org/10.1037/1076-898X.11.4.237 (2005).
38. Schmeck, A., Opfermann, M., van Gog, T., Paas, F. & Leutner, D. Measuring cognitive load with subjective rating scales during problem solving: differences between immediate and delayed ratings. *Instructional Science* **43**, 93–114 (2015).
39. Naismith, L. M., Cheung, J. J., Ringsted, C. & Cavalcanti, R. B. Limitations of subjective cognitive load measures in simulation-based procedural training. *Medical education* **49**, 805–814 (2015).
40. Goldstein, D. S., Bentho, O., Park, M.-Y. & Sharabi, Y. LF power of heart rate variability is not a measure of cardiac sympathetic tone but may be a measure of modulation of cardiac autonomic outflows by baroreflexes. *Experimental physiology* **96**, 1255–1261, https://doi.org/10.1113/expphysiol.2010.056259 (2011).
41. Montano, N. *et al*. Heart rate variability explored in the frequency domain: a tool to investigate the link between heart and behavior. *Neuroscience & Biobehavioral Reviews* **33**, 71–80 (2009).
42. Berger, R. D. *et al*. Beat-to-beat QT interval variability: novel evidence for repolarization lability in ischemic and nonischemic dilated cardiomyopathy. *Circulation* **96**, 1557–1565 (1997).
43. Haigney, M. C. *et al*. Gender differences and risk of ventricular tachycardia or ventricular fibrillation. *Heart Rhythm* **6**, 180–186 (2009).
44. Durning, S. J. & Artino, A. R. Situativity theory: A perspective on how participants and the environment can interact: AMEE Guide no. 52. *Medical Teacher* **33**, 188–199, https://doi.org/10.3109/0142159X.2011.550965 (2011).

45. Dobson, C. P., Kim, A. & Haigney, M. QT variability index. *Progress in cardiovascular diseases* **56**, 186–194 (2013).
46. Sokal, R. & Rohlf, F. *Biometry: The Principles and Practice of Statistics in Biological Research 4th edn* (WH Freeman and Co. 2012).
47. Hjortskov, N. *et al*. The effect of mental stress on heart rate variability and blood pressure during computer work. *European journal of applied physiology* **92**, 84–89 (2004).
48. Porges, S. W. The polyvagal theory: new insights into adaptive reactions of the autonomic nervous system. *Cleveland Clinic journal of medicine* **76**, S86 (2009).
49. Porges, S. W. Neuroception: A subconscious system for detecting threats and safety. *Zero to Three (J)* **24**, 19–24 (2004).

Acknowledgements

This work was supported by JPC-1 Award # 65126 (to S.D.). S.S. was supported by 5T32HL007227 (NIH).

Author Contributions

S.S., M.H. and S.D. designed the study. H.L. helped with subject recruitment and data collection. S.S. analyzed the data. S.S., M.H. and S.D. wrote the first draft of the manuscript. S.S., M.H., E.M.B., J.v.M., L.S., A.A., A.B., T.R. and S.D. discussed the results, reviewed and edited the final manuscript.

Additional Information

Competing Interests: The authors declare no competing interests.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Running head: Assessment of Sleep, Burnout and Stress Patterns

Well-Being in a Cohort of Active Duty Military Physicians: An Assessment of Sleep Patterns, Burnout and Perceived Stress

Divya Ramani[a], MS, Abigail Konopasky[a], PhD, Steven J. Durning[a], MD, PhD, Anthony R. Artino[a], Jr., MSC USN, and Jeffrey Mikita[b], MC USA

[a]*Graduate Programs in Health Professions Education, Uniformed Services University, Bethesda, Maryland,* [b]*Medical Modernization and Simulation Division, Defense Health Agency, Bethesda, Maryland*

Correspondence should be addressed to:

Divya Ramani, MS

Uniformed Services University

Department of Medicine

Graduate Programs in Health Professions Education

4301 Jones Bridge Road

Bethesda, MD 20814

Telephone: 703-945-9839

Email: divya.ramani.ctr@usuhs.edu

Funding Source

This study was funded by a grant from the Congressionally Directed Medical Research Program (CDMRP) JPC-1 strand (# NH83382416), which had no role in the design, conduct, or reporting of this study.

Declaration of Interest

None

Ethical approval

Ethical approval for this study was granted in December, 2016 by the Institutional Review Board at the Uniformed Services University in Bethesda, Maryland (#MED-83-3824). Written informed consent was obtained from each participant.

Disclaimer

The views expressed in this paper are those of the authors and do not necessarily reflect the official position or policy of the US Government, Department of Defense, Department of the Navy, or the Uniformed Services University.

Acknowledgments

The authors would like to thank Alexis Battista, Megan Ohmer, Anna Howle and Marta Smith for help with data collection.

Running head: Assessment of Sleep, Burnout, and Stress Patterns

Keywords: clinical reasoning; context; situated cognition; cognitive load

Text: 2321 words

Structured Summary: 290 words

Structured Summary

Introduction: Physician well-being is crucial and has the potential to impact patient safety. Many physicians across different stages of their careers experience stress, burnout and/or decreased sleep. These factors can negatively affect physician well-being and performance and contribute to medical errors. The purpose of this study is to further understand physician well-being by examining a single cohort for patterns of sleep, burnout and perceived stress across gender, training level and specialty.

Materials and methods: A cohort of 32 practicing military physicians ranging from first-year residents to attendings continuously wore an actigraphy watch for a duration of at least 5 days to

capture baseline sleep patterns. On the last day of data collection, participants completed a self-reported assessment of their daytime sleepiness using the Epworth Sleepiness Scale (ESS), a two-item burnout scale adapted from the Maslach Burnout Inventory, and a 10-item perceived stress questionnaire. Data for the entire cohort were descriptively analyzed.

Results: The cohort averaged a mean sleep duration of 6.69 hours across the 5 days, with a maximum mean sleep duration of 7.90 hours, and minimum mean sleep duration of 5.69 hours per day. Analysis stratified by gender and level of training revealed an average sleep duration of at least 6 hours across these groups. Overall, the cohort reported low perceived stress levels, low daytime sleepiness, and low burnout.

Conclusion: The cohort of physicians examined in the present study did not show signs of significant sleep deprivation, feelings of perceived stress or burnout. This may be due to the structure of military training facilities that emphasize duty hour regulations. In addition, these findings may be related to the fact that military health professionals are salaried, as opposed to being on a fee-for-service schedule, and military facilities offer well-being programs.

Introduction

The literature on physician well-being can tell a fairly bleak story.[1–4] Physicians' inherently challenging jobs and social and family responsibilities can all contribute to diminished sleep, increased stress, job burnout and medical errors.[1–3,5–7] Yet these effects may not be homogenous across all physicians: prior studies reveal differences in sleep deprivation, stress, and burnout across both gender and level of training (e.g., intern, resident, attending), with women and those at lower levels of training experiencing poorer sleep and worse stress and burnout.[1,4,8–10] Hence, the purpose of this descriptive study is to specifically examine patterns of

sleep, perceived stress, and job burnout together among military physicians, further exploring if there are differences across gender and training level.

Sleep plays a vital role in ensuring optimal brain and body function,[11] but the National Sleep Foundation's 2018 poll revealed that only 10% of Americans prioritize sleep over other aspects of a healthy lifestyle.[12] Sleep deprivation arises when healthy adults average less than 5 hours of sleep a day,[13] with 5 hours of sleep nightly or two nights of total sleep deprivation over a week both causing cognitive impairment.[13,14] Lack of sleep can lead to negative moods, anxiety, depression, irritability, confusion, and fatigue.[15] Sleep deprivation is common among professionals with demanding work schedules such as those in health care.[16,17] A study conducted on internal medicine residents revealed that being on call reduces total sleep time and negatively affects emotional equilibrium.[18] In another study of 33 surgical residents, a mean sleep of 5.3 hours reduced mental effectiveness by 80%, induced a high level of fatigue, and increased the risk of medical errors.[19] In a national survey, first- and second year residents averaging 5 or less hours of sleep a day reported conflict with other professional staff, accidents, consumption of alcohol and medication to stay awake.[7] The US Army Surgeon General implemented the performance triad (P3) in which sleep is considered as one of the important components along with nutrition and activity to maintain a healthy lifestyle in the military.[20] Yet only about 25% of Army physicians in a 2015 survey adhered to the sleep tenet.[20] This study further explores sleep patterns by examining a cohort of physicians across services at a military medical hospital.

Stress, which is defined as a biological response caused by either an internal or external stimulus, hampers mental and physical well-being.[21] Workplace stress is a common phenomenon among physicians.[8,9,22] A study of residents determined that they experience stress a result of role

Commented [1]: This needs to be refined based on reviewer comment

Commented [2]: define it in terms of workplace stress ---

Commented [3R2]: Maybe get rid of burnout

ambiguity, role overload, sleep deprivation, increased workload, as well as a mismatch between job responsibilities, and the individual's knowledge and skill set in fulfilling the responsibility.[22] In a comparison among physician expertise level (intern, resident and attending) poor sleep quality and increased patient load was found to be a major contributor to work stress.[9] Stress levels also appear to differ by gender: female physicians reportedly experience varied and more level of stress, owing at least in part to social-career conflicts.[4,8,23] This study explores the prevalence of stress in a cohort of military physicians across these identified areas of gender and level of training.

Stress is also regarded as one of the chief contributors to physician burnout.[24] Job burnout is a syndrome characterized by depersonalization, emotional exhaustion, and a sense of low personal accomplishment.[25] It occurs at a higher rate in medicine than in the general population of the United States.[2] A literature review revealed that burnout is prevalent across all stages of physician's lives: medical school, residency as well as in practice.[26] The 2019 Medscape National Physician Burnout, Depression & Suicide Report revealed that 44% of physicians in the United States report being burnt out, with female physicians 28% more likely than males to experience burnout.[27] Amongst the different specialties, internal medicine (49%), family medicine (48%), and general surgery (46%) were some of the most burned out.[27] In a longitudinal study conducted among resident physicians, higher rate of burnout symptoms was found among female residents and in general medicine.[28,29] Similarly, female surgeons showed higher levels of burnout and signs of depression than male surgeons.[29] Burnout is an issue in military physicians as well, with a recent study estimating that 26% of physician faculty in military graduate medical education experience burnout.[30]

Commented [4]: Maybe we dont have burnout as a component given that the scale to measure it is not correct

Conducted as a part of larger clinical reasoning study, this research describes the patterns of sleep, perceived stress, and burnout among a cohort of practicing military physicians in three of the specialties prone to sleep deprivation and burnout: internal medicine, family medicine, and surgery.[27] Based on the above literature suggesting that female physicians may experience more stress and burnout than male physicians, we also explored whether these patterns differ by gender. Also, based concerns about sleep deprivation in residency,[31] we examine differences across level of training (intern, resident, and attending).

Methods

This descriptive study was conducted at Uniformed Service University of the Health Sciences (USU) and the Walter Reed National Military Medical Center (WRNMMC). Practicing physicians from primary care and surgery volunteered to participate in this study. Sleep data were collected using an actigraphy watch (Philips Respironics – Actigraphy Spectrum Plus), which participants wore for a minimum period of 5 days. At the end of this period (for participants who wore the watch longer than 5 days, we took only the last 5 days of data), as part of the larger study, participants completed the Epworth Sleepiness Scale, Perceived Stress Scale, and the two-item version of the Maslach Burnout Inventory. The data collected were analyzed quantitatively using SPSS (statistical package for social science), and the research protocol was approved by the institutional review boards of USU and WRNMMC (MED-83-3824).

Measurements

Actigraphy. In this study participants wore an activity-sleep monitoring watch (Philips Respironics -Actigraphy Spectrum Plus) for a minimum period of 5 days for 24 consecutive hours (except while showering/swimming). The activity and sleep data collected were analyzed using the Actiware Software (Philips Respironics), providing information on hours of sleep per

day, maximum, minimum and average sleep time. Participants' sleep patterns were studied as a part of their everyday routine without any manipulation. Following Veasey et al., we defined sleep deprivation as less than 5 hours of sleep a day.[13]

Epworth Sleepiness Scale. The Epworth Sleepiness Scale is a widely used subjective measure of daytime sleepiness in the field of medicine and medical education.[32,33] Using a 0 (never dozing) to 3 (high chance of dozing) response scale, participants rated their recent tendency to doze/fall asleep in eight different situations. The score ranges from 0 to 24, with a score of 10 and above indicating daytime sleepiness.

Perceived Stress and Burnout Measures. To understand the well-being of this cohort, participants completed a 10-item perceived stress scale (PSS) that used a five-point response scale from 0 (never) through 4 (very often).[34] Perceived stress is the frequency with which situations in one's life are appraised as stressful. The score ranges from 0 to 40, with scores of 13 or below indicating low stress, scores of 14 to 26 indicating moderate stress, and scores of 27 and above indicating high stress.

Participants also completed a two-item measure adapted from the Maslach Burnout Inventory (MBI) that has been shown to reliably measure physician burnout.[35,36] One of the two items pertains to emotional exhaustion ("I feel burned out from my work") and the other pertains to depersonalization ("I have become more callous towards people since I took this job"). The measure employs a seven-point Likert-type scale (from 0, "never," to 6, "every day"). Consistent with prior reports, we considered a total score of 12 as high burnout, 6 to 11 as moderately burned out, and 5 and below as low burnout.

Results

A total of 32 practicing military physicians from USU and WRNMMC participated in the study; 22 were males and 10 were females (see Table 1). The sample included 16 attendings, 7 residents, and 9 interns, 26 in primary care and 6 in surgery (see Table 1). See Appendix A for data on all participants.

As measured by the actigraphy watch, the cohort had a mean nightly sleep of 6.69 hours (*SD* = .66), mean minimum nightly sleep of 5.7 hours (*SD* = .97) and mean maximum nightly sleep of 7.9 hours (*SD* = 1.16; see Table 2) over the 5-day study period. These actigraphy measures do not indicate sleep deprivation. Across gender, both male and female physicians obtained more than 6 hours of sleep, with females getting slightly more sleep than males (Table 2). Finally, physicians at all levels of training appeared to obtain adequate sleep, with interns sleeping slightly more on average than residents who slept slightly more on average than attendings (see Table 3). Overall, only 15.6% (*n* = 5) of participants had mean sleep amounts of less than 6 hours: two attendings, two residents, and an intern, all male (see Table 3).

Similarly, physicians' self-reports on the Epworth Sleepiness Scale revealed little daytime sleepiness on average (*M* = 6.81, *SD* = 3.89). Across gender, 23% of males and 10% of females reported daytime sleepiness (see Table 3). While standard deviations were high, residents reported slightly higher daytime sleepiness than interns and attendings (mean of 8.7 on ESS for the former versus 6.6 and 6.1 for the latter; see Table 2).

Results from the PSS indicated that physicians generally experienced low stress, with a mean of 10.6 (*SD* = 6.3). However, gender appeared to be a factor, with only about 32% of males reporting moderate to high stress, compared to 60% of females (See Table 3). Across level of training, there was little difference, with 33.3 % of interns, 33% of residents, and 44% of attendings reporting moderate to high stress (See Table 3).

Finally, this cohort reported low burnout overall, with a cohort mean of 3.66 (*SD* = 2.91) on the MBI. However burnout, like stress, appeared to differ according to gender, with 40% of females and only 27% of males reporting moderate to high burnout (see Table 3). Notably, only 11% (*N* = 1) of interns reported moderate to high burnout, compared to 57% of residents and 31% of attendings.

Discussion

This study examined three components of well-being – sleep, stress, and burnout – and their patterns across gender and level of training in a cohort of 32 military physicians. Overall, this cohort appeared to get an adequate amount of sleep and indicated little burnout and stress, with some small, but not statistically significant, differences emerging across gender and levels of training. These results are encouraging and could be the result of a number of factors, For example, the participants in this study worked in an academic medical facility that strictly follows duty hour regulations and is highly "mission-focused." That is, military physicians may experience higher job satisfaction and the sense of being a part of something "bigger" than other physicians.[37] Moreover, these physicians are salaried and are not required to see a specified number of patients per day, as some civilian facilities require; a fact that perhaps functions to reduce stress and burnout. Another factor that may have contributed to these results is that participants were allow to select their study times, and tended to do so during times of reduced workload (e.g., during a "light" rotation). As such, these reduced workload periods may have allowed participants to get more sleep and perhaps feel less stressed and burned out. Notwithstanding the uniqueness of the cohort studied here, this work a window into the *possibilities* that can be achieved when the appropriate work conditions are in place. This is both in line with the National Sleep Foundation and an improvement over the 2015 survey of U.S.

Army physicians.[11,20]

It is important to note, however, that even in this small cohort of physicians, there were sub-groups who appeared more likely to fall below the recommended sleep amounts and/or who experienced stress and burnout. While female physicians in this cohort got more hours of sleep than males (contrary to the finding of other studies[38]) more of these women reported being moderately to highly stressed and burned out, which seems to align with the literature.[4,8,23,27,28] Moreover, the results by level of training indicate that, for this cohort, interns and residents got more sleep than attendings. This may be due, in part, to the fact that interns and residents were often given time off from clinic or the wards to participate in the study. Alternatively, it could have been due to the notion that attendings have no duty hour regulations and were more likely to either be heading on or returning from a deployment. Further research should explore these trends. Meanwhile, interns and attendings (compared to residents) reported higher levels of stress while residents (compared to interns and attendings) reported higher levels of burnout. Residents may have expressed higher levels of burnout due to working close to duty hour regulations. However, this is a hypothesis and should be further explored in larger studies.

This study has several important limitations. First, it was a single-institution study that examined a relatively small number of participants. Also, participants were likely to be on less time consuming rotations that would be typical among trainees and attending physicians. That said, a strength of this study was the study design, which required participants, of multiple specialties and levels of training, to wear the actigraphy watches continuously over a five-day period.

Conclusion

The importance of physician is well established. This study examined the different

components of well-being and their pattern across a cohort of 32 military physicians. In contrast to prior work conducted nationally, findings from this study revealed a near optimal level of sleep, as well as low levels of stress and burnout across participants. There were, however, some small, but not statistically significant, differences in sleep and well-being across gender and level of training. While it is reassuring that this cohort of physicians appeared to be well-rested, unstressed, and no burned out, there is a need to further examine these components in larger investigations across a more diverse population.

References

1. Raj KS. Well-being in residency: a systematic review. *J Grad Med Educ*. 2016;8(5):674-684.
2. Shanafelt TD, Boone S, Tan L, et al. Burnout and satisfaction with work-life balance among US physicians relative to the general US population. *Arch Intern Med*. 2012;172(18):1377-1385.
3. Wallace JE, Lemaire JB, Ghali WA. Physician wellness: a missing quality indicator. *Lancet*. 2009;374(9702):1714-1721.
4. Dimou FM, Eckelbarger D, Riall TS. Surgeon burnout: a systematic review. *J Am Coll Surg*. 2016;222(6):1230.
5. van Wietmarschen H, Tjaden B, van Vliet M, Battjes-Fries M, Jong M. Effects of mindfulness training on perceived stress, self-compassion, and self-reflection of primary care physicians: a mixed-methods study. *BJGP open*. 2018;2(4):bjgpopen18X101621.
6. Lockley SW, Barger LK, Ayas NT, Rothschild JM, Czeisler CA, Landrigan CP. Effects of health care provider work hours and sleep deprivation on safety and performance. *Jt Comm J Qual Patient Saf*. 2007;33(11):7-18.

7. Baldwin Jr DC, Daugherty SR. Sleep deprivation and fatigue in residency training: results of a national survey of first-and second-year residents. *Sleep*. 2004;27(2):217-223.

8. Whitley TW, Gallery ME, Allison Jr EJ, Revicki DA. Factors associated with stress among emergency medicine residents. *Ann Emerg Med*. 1989;18(11):1157-1161.

9. Stucky ER, Dresselhaus TR, Dollarhide A, et al. Intern to attending: assessing stress among physicians. *Acad Med*. 2009;84(2):251-257.

10. Kassam A, Horton J, Shoimer I, Patten S. Predictors of well-being in resident physicians: a descriptive and psychometric study. *J Grad Med Educ*. 2015;7(1):70-74.

11. Czeisler CA. Duration, timing and quality of sleep are each vital for health, performance and safety. *Sleep Heal J Natl Sleep Found*. 2015;1(1):5-8.

12. 2018 Sleep in America® Poll – Sleep Prioritization and Personal Effectiveness. *Sleep Heal*. March 2018. doi:10.1016/J.SLEH.2018.02.007

13. Veasey S, Rosen R, Barzansky B, Rosen I, Owens J. Sleep loss and fatigue in residency training: a reappraisal. *Jama*. 2002;288(9):1116-1124.

14. Owens JA. Sleep loss and fatigue in medical training. *Curr Opin Pulm Med*. 2001;7(6):411-418.

15. Durmer JS, Dinges DF. Neurocognitive consequences of sleep deprivation. In: *Seminars in Neurology*. Vol 25. Copyright© 2005 by Thieme Medical Publishers, Inc., 333 Seventh Avenue, New …; 2005:117-129.

16. Weinger MB, Ancoli-Israel S. Sleep deprivation and clinical performance. *Jama*. 2002;287(8):955-957.

17. Alhola P, Polo-Kantola P. Sleep deprivation: Impact on cognitive performance. *Neuropsychiatr Dis Treat*. 2007.

18. Rose M, Manser T, Ware JC. Effects of call on sleep and mood in internal medicine residents. *Behav Sleep Med*. 2008;6(2):75-88.

19. McCormick F, Kadzielski J, Landrigan CP, Evans B, Herndon JH, Rubash HE. Surgeon fatigue: a prospective analysis of the incidence, risk, and intervals of predicted fatigue-related impairment in residents. *Arch Surg*. 2012;147(5):430-435.

20. Hsu DP, Hansen SL, Roberts TA, Murray CK, Mysliwiec V. Predictors of Wellness Behaviors in US Army Physicians. *Mil Med*. 2018;183(11-12):e641-e648.

21. Yaribeygi H, Panahi Y, Sahraei H, Johnston TP, Sahebkar A. The impact of stress on body function: A review. *EXCLI J*. 2017;16:1057.

22. Ebrahimi S, Kargar Z. Occupational stress among medical residents in educational hospitals. *Ann Occup Environ Med*. 2018;30(1):51.

23. McMurray JE, Linzer M, Konrad TR, et al. The work lives of women physicians. *J Gen Intern Med*. 2000;15(6):372-380.

24. Patel RS, Bachu R, Adikey A, Malik M, Shah M. Factors related to physician burnout and its consequences: a review. *Behav Sci (Basel)*. 2018;8(11):98.

25. Maslach C, Schaufeli WB, Leiter MP. Job burnout. *Annu Rev Psychol*. 2001;52(1):397-422.

26. IsHak WW, Lederer S, Mandili C, et al. Burnout during residency training: a literature review. *J Grad Med Educ*. 2009;1(2):236-242.

27. Kane L. Medscape national physician burnout, depression & suicide report 2019. *Available online at: medscape com/slideshow/2019-lifestyle-burnout-depression-6011056*. 2019.

28. Dyrbye LN, Burke SE, Hardeman RR, et al. Association of clinical specialty with

symptoms of burnout and career choice regret among US resident physicians. *Jama*. 2018;320(11):1114-1130.

29. Dyrbye LN, Shanafelt TD, Balch CM, Satele D, Sloan J, Freischlag J. Relationship between work-home conflicts and burnout among American surgeons: a comparison by sex. *Arch Surg*. 2011;146(2):211-217.
30. Summers SM, Nagy CJ, April MD, Kuiper BW, Rodriguez RG, Jones WS. The Prevalence of Faculty Physician Burnout in Military Graduate Medical Education Training Programs: A Cross-Sectional Study of Academic Physicians in the United States Department of Defense. *Mil Med*. 2019.
31. Commission J. Health care worker fatigue and patient safety. *Sentin event alert*. 2011;48(48):1-4.
32. Durning SJ, Capaldi II VF, Artino Jr AR, et al. A pilot study exploring the relationship between internists' self-reported sleepiness, performance on multiple-choice exam items and prefrontal cortex activity. *Med Teach*. 2014;36(5):434-440.
33. Johns MW. A new method for measuring daytime sleepiness: the Epworth sleepiness scale. *Sleep*. 1991;14(6):540-545.
34. Cohen S, Kamarck T, Mermelstein R. A global measure of perceived stress. *J Health Soc Behav*. 1983:385-396.
35. West CP, Dyrbye LN, Sloan JA, Shanafelt TD. Single item measures of emotional exhaustion and depersonalization are useful for assessing burnout in medical professionals. *J Gen Intern Med*. 2009;24(12):1318.
36. Durning SJ, Costanzo M, Artino Jr AR, et al. Functional neuroimaging correlates of burnout among internal medicine residents and faculty members. *Front psychiatry*.

2013;4:131.

37. Rus M, Sandu M. Personality factors and job satisfaction levels in military staff employed on a contract basis. *Procedia-social Behav Sci*. 2013;92:814-819.

38. Rodríiguez-Muñoz A, Moreno-Jimenez B, Fernandez-Mendoza JJ, Olavarrieta-Bernardino S, Vela-Bueno A. Insomnia and quality of sleep among primary care physicians: a gender perspective. *Rev Neurol*. 2008;47(3):119-123.

Table 1. Demographic information

	Intern	Resident	Attending
Gender			
Male	4	5	13
Female	5	2	3
Specialty			
Primary care	6	6	14
Surgery	3	1	2

Table 2. Comparison of study variable means across gender and level of training

	Minimum sleep *M* (*SD*)	Maximum sleep *M* (*SD*)	Average sleep *M* (*SD*)	Perceived sleepiness *M* (*SD*)	Perceived Stress *M* (*SD*)	Burnout *M* (*SD*)
Gender						
Male	5.5 (.85)	7.5 (1.03)	6.5 (.60)	7.2 (4.39)	9.6 (6.6)	3.36 (3.06)
Female	6.1 (1.14)	8.6 (1.17)	7.1 (.61)	5.9 (2.42)	12.5 (5.31)	4.3 (2.58)
Level of Training						
Intern	6.7 (4.27)	8.5 (1.3)	6.9 (.66)	6.6 (4.27)	8.7 (6.43)	2.4 (1.74)
Resident	5.8 (1.28)	7.6 (.82)	6.7 (.85)	8.7 (5.18)	12 (2.76)	5.4 (2.44)
Attending	5.5 (.79)	7.6 (1.09)	6.5 (.57)	6.1 (2.93)	10.9 (7.33)	3.6 (3.34)

Table 3. Percentage of participants experiencing sleep deprivation, daytime sleepiness, stress, and burnout

	Gender		Training Level		
	Male *N* =22	Female *N* =10	Intern *N* = 9	Resident *N* = 7	Attending *N* =16
Sleep deprivation (< 5 hours for 2 nights)	5% (*N* = 1)	0	0	0	6% (*N* = 1)
Daytime sleepiness (≥ 10 on ESS)	23% (*N* = 5)	10% (*N* = 1)	22% (*N* = 2)	29% (*N* = 2)	13% (*N* = 2)
Moderate to high stress (≥ 14 on PSS)	32% (*N* =7)	60% (*N* = 6)	33% (*N* = 3)	43% (*N* = 3)	44% (*N* = 7)
Moderate to high burnout (≥ 6 on MBI)	27% (*N* = 6)	40% (*N* = 4)	11% (*N* = 1)	57% (*N* = 4)	31% (*N* = 5)

Appendix A: Sleep, stress, and burnout scores across all participants

ID #	Training level	Gender	Min. sleep	Max. Sleep	Avg. Sleep	Epworth score	Stress score	Burnout score
1	Attending	Male	4.72	6.84	5.68	4	4.00	1
2	Resident	Male	5.07	6.28	5.69	5	16.00	3
3	Intern	Male	4.75	6.44	5.74	4	2.00	1
4	Attending	Male	3.63	7.23	5.77	4	10.00	1
5	Resident	Male	3.98	6.98	5.86	7	9.00	6
6	Attending	Male	5.55	6.28	6.00	4	3.00	1
7	Attending	Male	5.38	7.28	6.12	8	16.00	6
8	Attending	Male	5.31	6.88	6.12	7	20.00	2
9	Attending	Male	5.67	6.41	6.19	9	7.00	0
10	Attending	Male	5.38	8.23	6.23	5	27.00	10
11	Attending	Female	5.15	7.47	6.25	5	17.00	2
12	Intern	Male	4.78	7.75	6.41	16	7.00	1
13	Resident	Male	5.37	7.93	6.48	15	9.00	6
14	Attending	Male	5.20	7.97	6.52	12	7.00	4
15	Intern	Female	3.96	10.31	6.53	8	9.00	1
16	Attending	Male	5.84	7.22	6.62	5	3.00	0
17	Intern	Female	5.10	7.85	6.63	6	3.00	2
18	Attending	Female	6.17	7.46	6.84	5	18.00	7
19	Attending	Male	6.70	8.12	6.91	3	3.00	1
20	Resident	Female	5.97	8.60	6.96	5	10.00	3
21	Resident	Male	5.96	7.40	6.99	17	14.00	4
22	Attending	Female	6.22	7.79	7.07	9	14.00	9
23	Intern	Male	6.48	8.04	7.11	6	2.00	1
24	Resident	Male	6.80	8.05	7.13	8	12.00	10
25	Attending	Male	5.18	11.03	7.19	4	3.00	2
26	Intern	Male	6.06	8.25	7.31	3	16.00	3
27	Intern	Female	6.79	7.91	7.35	2	6.00	4
28	Attending	Male	6.63	8.29	7.45	2	14.00	8
29	Attending	Male	6.68	8.10	7.50	11	9.00	3
30	Intern	Female	6.77	10.61	7.56	10	15.00	3
31	Intern	Female	6.89	9.64	7.850	5	19.00	6
32	Resident	Female	8.02	8.41	8.20	4	14.00	6

ABSTRACT

Introduction: Contextual factors (e.g., diagnostic suggestion, burnout) can affect physician clinical reasoning performance, leading to diagnostic error. Yet, contextual factors have only recently been studied and none of that work focused on how physicians *appraise* (i.e., evaluate) the clinical situation as they reason. The purpose of this qualitative study was to use appraisal to describe the effect of contextual factors on clinical reasoning.

Materials and Methods: Physicians (n = 25) either viewed two video cases or participated in two live scenarios, one with contextual factors and one without. Afterwards, they completed a "think-aloud" reflection while re-viewing the cases. Transcribed think-alouds were coded for appraisal markers, comparing cases with and without contextual factors.

Results: When contextual factors are present, participants expressed more emotional evaluation and uncertainty about those emotions. Across all types of cases, participants expressed uncertainty about the case and assessed what "could" or "would" have gone differently.

Conclusions: This study suggests that one major effect of context specificity may be that it induces emotions which may affect the process of clinical reasoning and diagnostic error. It also suggests uncertainty may be common in clinical practice, and we should thus further explore its impact.

INTRODUCTION

As this physician watches herself on video and reasons aloud, she reflects on the standardized patient's concern about his chest pain and his desire for it not to be "scary." Thus, in the midst of determining diagnosis and treatment, she also must process this additional element in the situation, something to which she reacts with emotion ("concerned"), uncertainty ("possibly"),

and a proposition about what she *would have done* under different circumstances. In all these ways, she is *appraising* the clinical scenario: evaluating herself, others, and the situation.[1] This paper examines how physicians negotiate these contextual factors—features other than the content needed to arrive at the diagnosis (e.g., diagnostic suggestion in the quote above)–which can influence clinical reasoning through the lens of their appraisals.

Clinical reasoning has been described as the process ‘‘that enables practitioners to take wise action, meaning to take the best justified action in a specific context.”[2] This description acknowledges the vast complexity of clinical reasoning tasks, which include both the action taken (e.g., assigning a diagnosis, ordering a test, making a treatment plan) and the context for that action (e.g., a rushed appointment with a new patient who might also struggle with English). We approach the clinical context (including the participants, their interactions, and salient features of the environment) through the lens of *situated cognition*. This theory argues that individual cognition (reasoning in this case) is inseparable from the context in which it happens. From this theoretical perspective, situational (i.e., contextual) factors associated with the physician, patient, and encounter all interact (Figure 1).[3,4] Recent research indicates that these contextual factors can negatively impact the diagnostic process, resulting in *context specificity*: a physician arriving at two *different* diagnoses for two patients with the same diagnosis who also have the *same* symptoms and findings but different situations.[5,6] Context specificity undoubtedly results in diagnostic error, a problem that has come to the forefront of healthcare in the United States recently, affecting most of us at least once over the course of our lifetimes.[7]

Context specificity is recognized as an important problem in medical education,[8] but has only recently been studied, with emerging work suggesting that contextual factors (e.g., diagnostic suggestion by the patient, patient language difficulties, physician burnout, short appointment times) do affect performance across levels of experience, from residents to attending physicians.[3,9] In the military medical environment, where physicians often practice in austere settings with limited resources, it is arguably even more important to understand, and therefore mitigate, the effect of context specificity to reduce diagnostic error. We argue that a useful tool for understanding this effect is appraisal: explicit evaluations of the situation, either through *emotional evaluations* of self, patient, or environment (e.g., feeling "anxious"), modal verbs *assessing* what the physician might have done or didn't do (e.g., what the participant "would" have done differently), and hedging a claim to express *uncertainty* (e.g., noting that it is "most likely" [i.e., not definitely] gastroesophageal reflux disease (GERD); see the results section and Table 2 for more examples and explication).[1] Analysis of appraisal in medical education is limited, but has proven to be a useful lens for understanding how clinicians evaluate their own and others' roles in the clinical enterprise.[10,11] We discuss below how appraisal may interact with clinical reasoning, particularly in the presence of contextual factors.

How Emotional Evaluation, Assessment, and Uncertainty May Impact Clinical Reasoning

Emotional evaluations of the clinical encounter (e.g., emotional states like feeling "frustrated"), while not a primary focus in clinical reasoning literature, can offer telling appraisals of the clinical encounter, guiding as well as impeding the reasoning process.[12–15] Both positive and negative emotions can impact the reasoning process.[16] Moreover, negative emotional reactions

like anger, sadness, and shame can narrow one's attention, negatively affect risk estimation, lead to withdrawal and lowered empathy, and influence diagnostic accuracy through increased cognitive load.[12,14,17,18] Moreover, these effects can be made worse with the addition of other contextual factors like sleep deprivation and time constraints,[5,14] both of which are common in the deployed environment. Finally, a previous study examining mentions of contextual factors in case reflections found that emotional reactions were common and were usually associated with some form of tension, which the authors hypothesize may negatively affect reasoning.[6]

Another type of appraisal that can be associated with clinical reasoning is that of *assessment*, particularly self-assessment.[19] While different fields and lines of work operationalize assessment and self-assessment differently, we focus here on the broader pragmatic definition of assessment as an evaluation of self, other, or situation using modal verbs like "would" and negative markers like "not."[1] Informal assessments of clinical practice, like the ones studied here, may take the form of modal "would" or "could" statements about how physicians would optimize the relationship with the patient (e.g., engaging in conversation to put a patient at ease) or what additional information (e.g., labs, patient history) physicians could obtain in order to come to a diagnosis. This focus on what could be done differently offers an opportunity for reflection. Yet, as argued by McBee and colleagues, a desire for more information is closely related to difficulty with diagnostic closure.[6] In that study those who wanted more information had more difficulty with closure and, vice versa, those who were having difficulty with closure asked for more information.

The final type of appraisal in clinical reasoning that we focus on is that of *uncertainty*, which we broadly define as using hedgers like "kind of" or "possibly" to be less definitive about one's claims. While uncertainty can be beneficial (i.e., working against overconfidence, which can be associated with diagnostic error),[20] a body of research suggests it can negatively affect clinical reasoning outcomes and the stress level of the patient, and can lead to unnecessary testing and misspent time and money.[21,22] There is a growing recognition of the need to identify and manage uncertainty effectively,[21,23] but there is still limited research conceptualizing uncertainty in clinical reasoning, particularly regarding the phenomenon of context specificity.

The purpose of this qualitative study was to use markers of appraisal to describe the effect of contextual factors (e.g., diagnostic suggestion, non-native speaking patient) on the clinical reasoning process. We examine transcripts of physicians' oral reflections on clinical cases, asking:

1. How do participants use appraisals to refer to contextual factors, if at all?
2. Do participants appraise themselves and others differently in the presence of contextual factors, if at all? If so, how?

MATERIALS AND METHODS

As part of a larger study conducted at Uniformed Services University of the Health Sciences (USU) and Walter Reed National Military Medical Center (WRNMMC) examining the effects of contextual factors on clinical reasoning, practicing physicians in internal medicine, family

medicine, and surgery were invited via email to participate. After informed consent was obtained, they were quasi-randomly assigned (based on participant schedules) to one of three groups (video one, video two, or live scenario), each with a control (non-contextual) and contextual factor case (Table 1).

Participants in each of the two the video condition watched two cases, one with contextual factors and one without while participants in the live scenario condition participated in two live simulations, one with contextual factors and one without. All cases were straightforward depictions of common presentations in practice so that we could explore the impact of contextual factors on performance. Immediately after either watching each video case or participating in each live scenario case, participants were instructed to "think out loud" about their thoughts as they came to a diagnosis with no cuing or interruptions from the interviewer while either rewatching the short videos (video condition) or watching their own live simulations (live condition; see Battista et al. for a more detailed description of the procedure).[24,25] This "think-aloud" methodology has been shown in other studies to be an effective way to examine clinical reasoning.[3,26,27] Think-alouds lasted between 4 and 19 minutes and were then transcribed verbatim for analysis. The USU Internal Review Board granted approval for this research protocol (MED-83-3824).

Data analysis followed a four-step process. First, following our prior work on coding for context specificity,[6] we identified places in the think-alouds where participants explicitly mentioned the designated contextual factor (e.g., noting that the non-native speaker might not understand an English medical term). Second, we reviewed and discussed these examples as a coding team

(AK, DR, & MO), identifying markers of appraisal (drawn from systemic functional linguistics)[1,28] that would help us to track the effect of contextual factors in the transcripts.

Third, we selected three common appraisal markers that occurred frequently (450 times across the 50 transcripts) and aligned with prior work on clinical reasoning: emotional evaluations (e.g., mental states like *frustrated* and evaluations like *rude*), assessments (e.g., modal markers like *would* and negative markers like *not*), and hedgers denoting uncertainty (e.g., *kind of, possibly*; see Bhise et al. for a similar coding structure).[29-31] Finally, we compared the use of these markers in cases with and without contextual factors as an interpretive team.

RESULTS

Participants were 25 internal medicine (n = 18), family medicine (n = 2) and general surgery (n = 5) active-duty military physicians at USU and WRNMMC. Ten participants were female and 15 were male, and 15 were residents and 10 were staff physicians. Their ages ranged from 27 to 61 years old (m = 36 years old) with 0 to 34 years in practice (m = 7 years).

Of the 25 participants, 23 (92%) referred explicitly at some point to one of the designated contextual factors (i.e., diagnostic suggestion, circuitous history, non-native speaking patient, or questioning of physician's credentials). Most of the participants referred to the contextual factors at some point using appraisals: emotional evaluations, assessments, or hedgers of uncertainty (Table 2). In these examples, participants evaluated their own and the patient's emotions, primarily negative; they assessed their own and the video doctors' decisions, positing different or

additional actions that could have been taken; and they express uncertainty, both about the diagnostic process and their interpretation of the situation.

Our appraisal analysis revealed one notable difference between the types of cases: when reflecting on cases *with* contextual factors, participants offered more emotional evaluation, about three times per think-aloud without contextual factors and eight times per think-aloud with contextual factors (a *t*-test showed this to be a significant difference at the $p < .001$ level). In these emotional evaluations, they expressed their worry for the patient (particularly with the angina case, where several participants mention their "concern" about the patient's symptoms) or anger at the physician in the video (particularly with the diabetes case in which the patient struggles with English and the doctor does not bring in an interpreter; e.g., the fourth example, "callous," in Table 2). These emotional evaluations were usually negative (which is consistent with our design of the contextual factors which were typical distractors in practice), often offering evaluations of the *patient's* emotional state (e.g., the last example in Table 2). The cases without contextual factors, however, stimulated very few emotional evaluations.

Assessments, in contrast, appeared relatively equally both in cases with contextual factors (52 times across the sample) and without contextual factors (49 times across the sample). Participants used verbs like *would* or *should* and negation (e.g., *not*, *didn't*) to make assessments about what did not happen and probably should have. In other words, they made statements about some *optimal* condition in which they had more information or behaved differently. These assessments were primarily about what should have been asked in the patient history, what should have been done on physical exam, and what tests should have been performed for

management. For example, in referring to a video of an angina case without contextual factors, a participant says, "I **would have asked** him if he has any nausea when he's having this pain. I **would have asked** him whether or not he had any diaphoresis, or sweating, when he had the pain. She [physician in the video] **didn't really ask** about any of those kinds of things, so I think those all would have been helpful." Here the participant assessed what the physician in the video did *not* do and what he *would have* done in the same situation. Participants made these same sorts of assessments about themselves, most often in the live scenario cases, as in this comment on the diabetes case without contextual factors: "I **didn't actually order** a thyroid, now that I think about that [...] But that would be something I **would want to get**, too."

Additionally, we found that the assessments and emotional evaluations discussed above were combined with hedgers of uncertainty. For example, in discussing the angina case with contextual factors, a participant used both modal verbs of assessment (bold) and hedgers of uncertainty (underlined), "I **would** <u>probably</u> **trial** him on a PPI [proton pump inhibitor], um <u>kind of</u> counsel him to avoid foods that trigger GERD. [...] I think that's <u>probably</u> what I **would** <u>probably</u> **do** for him." Thus, even in discussing optimal cases of what one would or should do, participants tended to express hedging towards uncertainty. This type of appraisal, like assessments, occurred across both conditions (e.g., it's **probably, not necessarily** related here"), with 146 hedgers coded for cases without contextual factors and 151 hedgers coded for cases with contextual factors. As we might perhaps expect in a think-aloud task asking for thoughts toward a diagnosis, most of the hedging related in some way to the process of coming to a diagnosis, particularly connecting findings from the history with the diagnosis (either leading or differential), as in Table 3. In these examples, participants were denoting uncertainty in the

posited connections among evidence (e.g., smoking history) and possible diagnoses (e.g., angina). Note that in the third example, the participant was explicitly moderating the force of his *own reasoning*, noting that he was *kind of done*.

While examples like those in Table 3 occurred similarly across conditions, we also found some types of uncertainty hedging that occurred most frequently in the presence of contextual factors, related to: (a) participant affect, (b) patient affect, (c) patient characteristics, and (d) assessment of patient credibility, shown in Table 4. In the first two examples, participants softened references to *feelings*--their own or their patient's, minimizing the sense of concern. In the last two examples, participants softened references to patients that might not be interpreted as kind or complimentary. In all these examples, the contextual factor brought along with it some kind of emotion or judgment, so the participant took the conversational time and energy to soften the statement. One consequence of this is that participants hedged statements around diagnosis and clinical reasoning (Table 3) *less*, focusing their conversational time and energy on issues related to the contextual factors.

DISCUSSION

This study offers important insights into the *process* of clinical reasoning as reflected in the think-alouds of these active-duty physicians and how that process differs in the presence of contextual factors. First, explicit mentions of contextual factors are usually accompanied by some kind of appraisal, whether it is evaluating the emotions at play (one's own or the patient's), assessing the clinical actions and diagnostic steps taken, or hedging to express uncertainty about

the diagnostic process. Second, in cases with contextual factors, participants do more emotional evaluation and hedging of emotions or judgments. Thus, the presence of these contextual factors seems to stimulate participants to reflect more on emotional states and to spend their reflection time carefully qualifying (i.e., hedging) their claims about emotions and patients. Third, whether or not contextual factors are present, physicians regularly assess in their reflections how a clinical encounter "could" or "would" have gone and express uncertainty about what is "probably" or "possibly" going on clinically. In the context of these think-aloud reflections, assessment of what has been done and uncertainty about what is going on seem to be more the norm than the exception.

This work has teaching implications, particularly around emotions and their role in clinical education and practice. Recent research with internal medicine residents suggests that, when unexamined, negative emotions like shame can fester, leading to emotional distress, impaired self-regulation, and disengagement from learning, among many other negative outcomes.[32] If negative emotional evaluations are indeed more prominent in the presence of contextual factors, medical educators should be aware of this and the additional support learners might need in these situations, both regarding the clinical reasoning process and their own and the patient's emotional health. This is particularly important, since physician and patient emotions heavily affect each other and the diagnostic process.[13] Currently, medical school curricula are focused far more on content than on context; perhaps more attention to context—and how to notice and manage the emotions and mitigate uncertainty—would lead to better emotional regulation in the face of these contextual factors.

These findings suggest several implications for research in clinical reasoning. To begin, if contextual factors somehow steer physicians to spend more reflective time on emotional evaluations, particularly negative ones, it is critical to better understand the effects of these evaluations on clinical reasoning. For instance, research in the appraisal-tendency framework suggests that not all negative emotions have the same effect on reasoning. For instance, while anger and sadness are both negative emotions, anger leads people to attribute control to individuals (e.g., anger at the individual doctor in one of the study videos for not calling in an interpreter) and to be more certain about their judgments.[16,33] Meanwhile, sadness leads people to attribute control more generally to the situation (e.g., sadness for the situation in which our patient with angina tries desperately to convince the physician that he has GERD, a less "scary" situation) and to be less certain about their judgments. In addition, while some negative emotions, like shame or disappointment, tend to be psychologically "deactivating" for physicians, thereby resulting in disengagement from a given situation, other negative emotions, like confusion or frustration, can actually lead to greater arousal and situational engagement (and, potentially, enhanced performance.[34] Future research could explore how eliciting specific types of negative and positive emotions (both activating and deactivating) affect diagnostic certainty, particularly anchoring and diagnostic closure.[6,34]

Next, participants across cases with and without contextual factors regularly express alternate paths and uncertainty, assessing how they "could" or "would" proceed combined with the hedging of what they "might" or "probably" think is happening. This suggests that uncertainty is quite common in clinical practice. Indeed, it may be more common than the literature suggests as our cases were depicted as being straightforward and common to practice and included

practicing attendings and residents. Nonetheless, uncertainty was still a common theme, perhaps indicating that we should further explore the impact of uncertainty on clinical reasoning and diagnostic error. McBee and colleagues propose, for instance, that uncertainty can hinder diagnostic closure,[6] which could generate unnecessary costs and delay care. This could be a particular problem in acute and austere military environments where there may be few additional diagnostic procedures one could initiate in an effort to alleviate uncertainty before moving forward with treatment. Studies further exploring how an "it totally possibly could be" attitude affects the diagnostic process and diagnostic outcomes are needed.

A final research implication is methodological: appraisal, drawn from the functional linguistics tradition,[1,28] appears to offer effective tools for studying clinical reasoning. This is particularly true when taking a situated cognition perspective, in which researchers seek to explain how a variety of factors (related to physician, patient, and situation) interact moment-by-moment in the clinical reasoning process.[3] As previous work suggests, appraisals offer a rich window into the practice of medicine and medical teaching, revealing the varying stances clinicians take towards themselves, their patients, and the broader clinical and institutional environment.[10,11]

There are several limitations to this study. First, all participants were drawn from only two (closely related) sites, so these patterns may not hold for other military physicians. Future work should explore not only other sites but other types of physicians (e.g., critical care, oncology) to offer a broader picture. Second, think-aloud reflections were retrospective (i.e., after participating in or viewing the case and giving the diagnosis), but this allowed participants to concentrate fully on the case as they were coming to their diagnosis without having to verbalize.

Meanwhile, prior work has shown the value of retrospective think-alouds for examining clinical reasoning.[3,35] Finally, we drew this data set from both live and video simulation modalities, which are different experiences for the participants. Because we were finding similar results across the two modalities and because our focus was on the contextual factors, we chose to combine live and video. Future planned analysis in this project will further explore the differences between those modalities.

CONCLUSIONS

Given the relationship between contextual factors and diagnostic error,[36] continued qualitative work like this exploring how physicians reason through contextual factors moment-by-moment is critical. This study suggests that one major effect of context specificity may be that it induces emotions, negative ones in these cases, that might affect processing speed or attention.[3] Perhaps if we help physicians to recognize these emotions and their potential effect on clinical reasoning, we might have a positive impact on diagnostic error. Moreover, this study highlights the prevalence of uncertainty even in straightforward cases *without* contextual factors. Future work could examine how physician contextual factors like sleep deprivation and burnout might affect diagnostic reasoning in the presence of uncertainty, even when context specificity is not at play.[9,37,38] The language physicians use to reflect upon their practice—i.e., not only *what* they say but *how* they say it--can offer valuable insight into how they make assessments and evaluations about themselves, their patients, and the broader clinical situation.

REFERENCES

1. Martin JR, Rose D: Working with Discourse: Meaning beyond the Clause. Bloomsbury Publishing, London, 2003.
2. Higgs J, Jones MA, Loftus S, Christensen N: Clinical Reasoning in the Health Professions. Philadelphia, PA, Elsevier Health Sciences, 2018.
3. Durning SJ, Artino AR, Boulet JR, Dorrance K, van der Vleuten C, Schuwirth L: The impact of selected contextual factors on experts' clinical reasoning performance (does context impact clinical reasoning performance in experts?). Adv Heal Sci Educ 2012; 17(1): 65-79.
4. Brown JS, Collins A, Duguid P: Situated cognition and the culture of learning. Educ Res 1989; 18(1): 32-42.
5. Durning S, Artino Jr AR, Pangaro L, van der Vleuten CPM, Schuwirth L: Context and clinical reasoning: understanding the perspective of the expert's voice. Med Educ 2011; 45(9): 927-38.
6. McBee E, Ratcliffe T, Picho K, et al.: Consequences of contextual factors on clinical reasoning in resident physicians. Adv Heal Sci Educ 2015; 20(5): 1225-36.
7. National Academies of Sciences and Medicine: Improving Diagnosis in Health Care. Washington, DC, National Academies Press, 2015.
8. ten Cate O, Durning SJ: Understanding clinical reasoning from multiple perspectives: A conceptual and theoretical overview. In: Principles and Practice of Case-based Clinical Reasoning Education. Edited by ten Cate O, Custers E, Durning S. Berlin, Germany, Springer International Publising, 2018: 35-46. doi:10.1007/978-3-319-64828-6_3
9. McBee E, Ratcliffe T, Picho K, et al.: Contextual factors and clinical reasoning: differences in diagnostic and therapeutic reasoning in board certified versus resident

physicians. BMC Med Educ: 2017;17(1):211.

10. Ferguson A: Appraisal in student–supervisor conferencing: A linguistic analysis. Int J Lang Commun Disord 2010; 45(2): 215-29.
11. Gallardo S, Ferrari L: How doctors view their health and professional practice: An appraisal analysis of medical discourse. J Pragmat 2010; 42(12): 3172-87.
12. Blanchette I, Richards A: The influence of affect on higher level cognition: A review of research on interpretation, judgement, decision making and reasoning. Cogn Emot 2010; 24(4): 561-95.
13. Langridge N, Roberts L, Pope C: The role of clinician emotion in clinical reasoning: Balancing the analytical process. Man Ther 2016; 21: 277-81.
14. Resnick ML: The effect of affect: Decision making in the emotional context of health care. In: 2012 Symposium on Human Factors and Ergonomics in Health Care; 2012: 39-44.
15. Stolper E, Van de Wiel M, Van Royen P, Van Bokhoven M, Van der Weijden T, Dinant GJ: Gut feelings as a third track in general practitioners' diagnostic reasoning. J Gen Intern Med 2011; 26(2): 197-203.
16. Lerner JS, Li Y, Valdesolo P, Kassam KS: Emotion and decision making. Annu Rev Psychol 2015; 66.
17. Bynum IV WE, Goodie JL: Shame, guilt, and the medical learner: ignored connections and why we should care. Med Educ 2014; 48(11): 1045-54.
18. Fraser K, Ma I, Teteris E, Baxter H, Wright B, McLaughlin K: Emotion, cognitive load and learning outcomes during simulation training. Med Educ 2012; 46(11): 1055-62.
19. Eva KW, Regehr G: Self-assessment in the health professions: a reformulation and

research agenda. Acad Med 2005; 80(Supplement): S46-54. doi:10.1097/00001888-200510001-00015

20. Berner ES, Graber ML: Overconfidence as a cause of diagnostic error in medicine. Am J Med 2008; 121(5): S2-23.
21. O'Riordan M, Aktürk Z, Ortiz JMB, et al.: Dealing with uncertainty in general practice: an essential skill for the general practitioner. Qual Prim Care 2011; 19(3).
22. Zaat JOM, van Eijk JTM: General practitioners' uncertainty, risk preference, and use of laboratory tests. Med Care 1992: 846-54.
23. Kennedy AG: Managing uncertainty in diagnostic practice. J Eval Clin Pract 2017; 23(5): 959-63.
24. Ericsson KA, Simon HA: How to study thinking in everyday life: Contrasting think-aloud protocols with descriptions and explanations of thinking. Mind, Cult Act 1998; 5(3): 178-86.
25. Battista A, Konopasky A, Ramani D, et al.: Clinical reasoning in the primary care setting: Two scenario-based simulations for residents and attendings. MedEdPORTAL 2018; 14(10773).
26. Burbach B, Barnason S, Thompson SA: Using "think aloud" to capture clinical reasoning during patient simulation. Int J Nurs Educ Scholarsh 2015; 12(1):1-7.
27. Funkesson KH, Anbäcken E-M, Ek A-C: Nurses' reasoning process during care planning taking pressure ulcer prevention as an example. A think-aloud study. Int J Nurs Stud 2007; 44(7): 1109-19.
28. Halliday MAK, Matthiessen C, Halliday: An Introduction to Functional Grammar. New York, NY, Routledge, 2014.

29. Makary MA, Daniel M: Medical error—the third leading cause of death in the US. Bmj 2016; 353: i2139.
30. Levinson P, Brown P, Levinson SC, Levinson SC: Politeness: Some Universals in Language Usage. Vol 4. New York, NY, Cambridge University Press, 1987.
31. Bhise V, Rajan SS, Sittig DF, Morgan RO, Chaudhary P, Singh H: Defining and measuring diagnostic uncertainty in medicine: a systematic review. J Gen Intern Med 2018: 1-13.
32. Bynum IV WE, Artino Jr AR, Uijtdehaage S, Webb AMB, Varpio L: Sentinel Emotional Events: The Nature, Triggers, and Effects of Shame Experiences in Medical Residents. Acad Med 2018; 94(1): 85-93.
33. Lerner JS, Keltner D: Beyond valence: Toward a model of emotion-specific influences on judgement and choice. Cogn Emot 2000; 14(4): 473-93.
34. Artino Jr AR, Holmboe ES, Durning SJ: Can achievement emotions be used to better understand motivation, learning, and performance in medical education? Med Teach 2012; 34(3): 240-4.
35. Durning SJ, Graner J, Artino AR Jr, et al.: Using functional neuroimaging combined with a think-aloud protocol to explore clinical reasoning expertise in internal medicine. J Mil Med 2012; 177(suppl_9): 72-8.
36. Trowbridge RL, Graber ML: Clinical reasoning and diagnostic error. In: Teaching Clinical Reasoning. Edited by Trowbridge RL, Rencic JJ, Durning SJ. Philadelphia, PA, American College of Physicians, 2015: 1-12.
37. Halbesleben JRB, Rathert C: Linking physician burnout and patient outcomes: exploring the dyadic relationship between physicians and patients. Health Care Manage Rev 2008;

33(1): 29-39.

38. Pilcher JJ, Huffcutt AI: Effects of sleep deprivation on performance: a meta-analysis. Sleep 1996; 19(4): 318-26.

The Linguistic Effects of Context Specificity: Exploring Affect, Cognitive Processes, and Agency in Physicians' Think-Aloud Reflections

Introduction

Diagnostic error is a national – if not international – crisis and is frequently cited as a leading cause of death in the United States.[1,2] One important source of error relates to context specificity, a phenomenon whereby a physician sees two patients with identical presentations and yet comes to two different diagnostic decisions.[3,4] Situated cognition theory, which argues that knowing cannot be separated from context, activity, or language,[5] provides a useful framework for exploring context specificity. In the case of clinical reasoning, one way to explore context specificity is to account for *contextual factors*: elements in the specific situation that pertain to the physician (e.g., burnout, sleepiness), patient (e.g., spoken language proficiency, challenging physician credentials), and environment (e.g., time for appointment, availability of ancillary staff), but that are not pieces of information traditionally viewed as needed to establish the diagnosis or management and often are unique to the situation. These contextual factors interact as the encounter unfolds and, from the situated cognition perspective, clinical reasoning emerges. Therefore, situated cognition provides a useful theoretical lens for exploring errors and other variation in physician performance that results from context specificity.

The literature suggests that affect, higher-level cognitive processes (e.g., problem solving and decision making[6]), and an individual's agency (the capacity to produce an effect[7,8]) are important for reasoning[8–17]; however, we do not know how these factors interact and respond in different contexts (i.e., situations). Moreover, only affect has been explored in the context of *clinical* reasoning.[9–11] Linguistics provides a mechanism for understanding how different

contexts may impact affect, cognitive processes, and agency, providing a potential means for better supporting physician performance in the presence of contextual factors as well as helping to unravel the vexing phenomenon of context specificity. Therefore, the purpose of this study is to use situated cognition theory and linguistic analysis to determine whether contextual factors lead to differences in affect, cognitive processes, and individual agency, and, if so, to describe these differences.

Affect, Cognitive Processes, and Agency in Clinical Reasoning

While much of the clinical reasoning literature focuses on the logical aspects of cognition, emotion (an affective state characterized by arousal that results from a specific stimulus in the environment[9]) is also an integral part of the reasoning process.[9,10,18,19] The greater the magnitude (increase in level of arousal) of the emotion, the greater the possible effect on the clinical reasoning process.[11] But some emotional arousal is present in *all* reasoning,[19] particularly in the high-stakes context of patient care, where anxiety and stress often exist.[9] Negative emotions like anxiety can cause a narrowing of attention and risk aversion (along with potentially resultant narrowing of cognitive capacity and increasing cognitive load), which, in turn, can increase the chance of medical error in the form of missed or delayed diagnoses.[11,20–22] Positive emotions can often support reasoning, but they can also lead to overconfidence which can, in turn, result in less information gathering during a patent encounter.[11,20,23] Recent research has also linked emotions to contextual factors, with study participants voicing primarily negative emotional reactions to various contextual factors.[24,25] One potential solution proposed by those interested in ameliorating diagnostic error is to increase *explicit awareness* of these emotions, taking them seriously and exploring how different contextual factors may trigger different

emotional states.[10,11] Linguistic analysis is one such path to increased awareness of emotions,[12,26] offering a novel way of exploring how various emotions are triggered by contextual factors.

While the major *outcomes* of clinical reasoning may be diagnostic and management plans, clinical reasoning itself is also a complex *process* of meaning making that scholars are only now beginning to fully understand.[25,27–29] Contained under the umbrella of the clinical reasoning process are narrower cognitive processes like problem representation, hypothesis generation, hypothesis testing, and metacognition (which involves, among other things, being aware of, controlling, and managing one's cognition in pursuit of a task[30]).[31,32] Sometimes these cognitive processes are conscious and sometimes they are unconscious,[32] and they appear to be inhibited by certain contextual factors.[24,28,33] The presence of these inhibiting contextual factors can increase cognitive load, defined as perceived mental effort.[33,34] When the cognitive load is too high for a clinician, their reasoning can be negatively affected, leading to diagnostic error.[33,35] In order to study cognitive processes in clinical reasoning and how they may be related to cognitive load, we can explore the patterns of distinct linguistic markers like *think*, *know*, or *consider*.[36] This allows us to examine whether and how expression of cognitive processes shifts under the influence of context specificity. Following Khawaja and colleagues, we predicted that higher cognitive load would be associated with more cognitive process markers as individuals worked to actively understand their situation.[36]

While situated and context-dependent, the process of clinical reasoning is largely *directed* by the physician. It is the physician who marshals the necessary resources – some of which may involve other people (e.g., specialty consultation) and diagnostic artifacts (e.g., diagnostic imaging, labs) – to eventually reason to a decision. Yet, as discussed above, physicians may feel uncertain or anxious in the presence of contextual factors, letting their emotions guide their

reasoning.[20,21,24] We approach this through the lens of *agency*, (broadly defined as the capacity to produce an effect[7,8]), exploring whether contextual factors affect how physicians talk about themselves as agents (or not) of the reasoning process. In particular, we examined the frequency of the first-person singular pronoun *I*, since it has been argued in prior work to be indicative of a feeling of individual, intentional causation, particularly in comparison to other pronouns like generic *you*.[8,15,37] Moreover, the first-person singular pronoun has been associated with greater depth of reflection in medical student essays[38] in one study and decreased cognitive load in team problem solving in another.[36] These studies along with the broader cognitive load literature suggest additional reasons why we might expect to see decreased *I*-usage in the presence of contextual factors.

In order to examine and describe potential effects of context specificity on affect, cognitive processes, and individual agency so that we can better support clinicians, we pose the following research questions:

1. Does the presence of contextual factors in cases lead to differences in linguistic measures of affect, cognitive processes, or individual agency?
2. If so, what are the patterns of different subtypes of affect, cognitive processes, and agency in cases with and without contextual factors?

Based on the literature reviewed above, we hypothesized that increased cognitive load in the condition with contextual factors would lead to a greater frequency of affect and cognitive process markers and a lower frequency of first-person singular pronouns.

Materials and Methods

This study is a comparative and exploratory linguistic analysis of think-aloud reflections drawn from a larger investigation[28,39] of context specificity and clinical reasoning at Uniformed Services University of the Health Sciences, Walter Reed National Military Medical Center, and Naval Medical Center San Diego. The study was approved by the institutional review boards at all three sites. Physicians in internal medicine, family medicine, and surgery were quasi-randomly assigned to a video or live scenario condition. Participants in the video condition viewed one patient encounter with contextual factors (e.g., low English proficiency, diagnostic suggestion) and one without. After viewing each case and determining the diagnosis and management plan (see Durning et al., 2014, for the format used[40]), participants were asked to immediately rewatch the video and "think aloud" about how they came to their diagnosis. Participants in the live scenario condition experienced the same cases, also one with contextual factors and one without, but participated in the case as a physician with a simulated participant as the patient rather than viewing a video. After giving a diagnosis and management plan in the same format as the video condition participants, they watched the encounter they had just participated in and immediately conducted a think-aloud procedure. Participants in both conditions worked with cases that had typical presentations of common diseases: diabetes mellitus and unstable angina. The case content was controlled (i.e., identical presenting symptoms, language and gestures to represent those symptoms, and physical findings); thus, the only differences between the cases with and without contextual factors were the contextual factors themselves.

Think-Aloud Procedure

For the think-aloud procedure, participants were asked to speak their thoughts out loud, without making judgments or offering insights, as they engaged with the task (e.g., a video of an event[41]). Past work has indicated that think-aloud transcripts represent a reasonable measure of thinking,[41–43] as well as an effective way to assess clinical reasoning.[44–46] In this study, participants were given brief instructions and a warm-up exercise in the think-aloud method prior to engaging in the cases. Then, after either viewing the video case or participating in the live scenario and determining the diagnosis and management, they were prompted to think aloud about their thoughts leading to diagnosis and treatment. Participants were given up to 30 minutes to complete this and were allowed to stop or rewind the video.

Data Analysis

To understand how the process of clinical reasoning is affected by context specificity, we used Linguistic Inquiry and Word Count (LIWC) software. LIWC is a transparent (i.e., coded words and phrases are accessible to researchers) text analysis program that codes for affect, cognitive processes, and agency, among other psychological processes.[12] We coded all transcripts with LIWC for the broad categories of affect and cognitive processes and the subcategory of first-person singular pronouns (i.e., *I* and *me*). To control for the potential effect of varying word counts, LIWC calculates a percentage of coded categories per 100 words (e.g., if there were 10 affect-related words in a 200-word transcript, LIWC assigns that transcript a value of 5% for affect). We then conducted a repeated measures multivariate analysis of variance (MANOVA) and follow-up univariate analyses with affect, cognitive process, and first-person

pronouns as the dependent variables, comparing participants' language in the cases with and without contextual factors.

To explore patterns in affect, cognition, and agency, we examined descriptive statistics of the subcategories making up affect (positive emotions, negative emotions, anxiety, anger, and sadness) and cognitive processes (insight, causal processes, certainty, tentativeness, discrepancy, and difference – described in greater detail in the results section below). We also examined descriptive statistics of other personal pronouns (*we*, *you*, *he/she*, and *they*) to better understand how individual actions interacted with the actions of others in these data.

Results

Participants were 64 internal medicine, family medicine, and surgery physicians; 22 were women and 41 were men (See Table 1 for demographic details). Think-aloud transcripts of cases without contextual factors (n = 64) were between 198 and 1903 words (m = 458) and those with contextual factors (n = 64) were between 256 and 2293 (m = 513). Across all transcripts, affective markers represented between 1.4% and 10.4% of the words, cognitive processing words between 9.9% and 25.5%, and first-person singular pronouns between 0.2% and 9.6%.

The Effects of Context Specificity: Affect, Cognitive Processing, and Individual Agency

Repeated measures MANOVA results revealed significant differences in cases with and without contextual factors (Pillai's Trace = .22, F = 5.6, df = [3, 61], $p < .01$). Follow-up univariate analyses indicated that participant language contained more affective and cognitive process markers in think alouds of cases with contextual factors. Additionally, in think alouds without contextual factors, participants used more first-person singular *I/me* pronouns,

suggesting a greater expression of individual agency (see Table 2). These differences were, however, statistically significant only for cognitive processing words, not for affective markers and *I/me* pronouns.

In order to better understand the differences in affect in cases with and without contextual factors, we explored the three subcategories comprising LIWC's affect category that emerged most frequently in our data: positive emotions, negative emotions, and anxiety (a subcategory of negative emotions; see Table 3). The difference between conditions resulted from more negative emotions in cases with contextual factors, where participants thought aloud about the standardized patient's emotions (stress, anxiety) and their own thought processes (e.g., thinking "that's ridiculous" about a potential diagnosis of coal worker's lung). LIWC also identified some medical terms (e.g., *stress* test, head *trauma*, *resolves* with rest), but these uses appear in both conditions (with and without contextual factors).

Next, LIWC's cognitive processes category derives from six subcategories: insight, causal processes, certainty, tentativeness, discrepancy, and difference (see Table 4). The greatest contrast (and the only statistically significant difference in our exploratory analysis) appears to be in terms of insight (terms associated with learning or understanding like *think, explain, evaluate,* or *consider*[12,47]): participants talked more about their learning or understanding when contextual factors were present, more often explicitly reflecting on their *thinking* or *considering*. While the other differences were not statistically significant, it is notable that, in the presence of contextual factors, participants seemed to use fewer markers of certainty (terms indicating a certain level of conviction like *clear, sure, certainly,* or *namely*[12]) and more markers of tentativeness (terms indicating a hedging or uncertain stance like *kind of, may, if,* or *anything*[12]). Similarly, participants made more discrepancies (terms indicating a difference between an actual

and possible state like *should, would, could,* and *need*[12]) in the presence of contextual factors, often conveying a speculation about what could or would be the case, given some condition (e.g., "Her HCTZ [dose] **could** be improved....[so] her lifestyle **could** improve."). Finally, markers of causal process (terms implying that one thing gives rise to another like *how, based, because,* or *why*[12]) and difference (terms of distinction, including negation, like *but, really, not,* or *other*[12]) appeared to be similar across conditions.

Finally, examining agency beyond first-person pronouns (discussed above), participants appeared to use third-person pronouns in the presence of contextual factors to focus on the actions of *others*, often the patient or, in the video cases, the doctor depicted in the video. In fact, thinking aloud about the actions of third-person singular others was the most common pronominal use across conditions (between 4.8% and 5.5% of the word count), but these exploratory analyses suggest that *he/she/him/her* usage goes up in the presence of contextual factors as *I/me* goes down. Despite the increase in cognitive processing words (which often have *I*-subjects) in the presence of contextual factors, the overall focus on the self's actions yielded to thinking aloud about the patient and video doctor, often with reference to a contextual factor. For example, here a participant reflects on the case with a patient who is not a native speaker of English: "**She** [patient] asks **him** [doctor] about speaking Spanish and **he** says **he** only speaks English." This participant only referred to herself eight times (0.9% of words) in this case, while she referred to herself 45 times (4.3% of words) in the non-contextual factors case.

Discussion

This study demonstrates how linguistic tools can offer insight into the situated nature of the clinical reasoning process: when contextual factors are present, participants verbalize their

cognitive processes more as they work to make sense of the situation and the case. Also, while not statistically significant, the trends in the hypothesized direction suggested that participants voice more emotions and fewer of their own thoughts and actions (as measured by first person pronouns) in the presence of contextual factors. These findings corroborate the predictions that emerge from situated cognition and cognitive load theory; specifically, that contextual factors would engender higher cognitive load and, thus, more cognitive processes and emotion and less focus on the self (versus the contextual factors themselves).

Moreover, descriptive findings from this study offer further insight into how participants react to context specificity. First, the major difference in affect markers was in negative emotions, and most of this negative affect centered on the simulated participant or the participant's own reasoning processes. While positive emotions can also affect reasoning,[11,20] negative emotions are more frequently associated with error, which is what we see in the presence of contextual factors. As with prior studies of emotion and clinical reasoning,[9,18] anxiety was common, even with physicians solving typical cases for their field. This suggests the need to be more mindful of the effects of contextual factors, including helping physicians identify and mitigate stress and anxiety during clinical encounters.

Second, the cognitive process marker that was most strongly associated with context specificity was LIWC's "insight" category, which is language associated with understanding. The presence of a contextual factor, then, appears to focus participants' verbalizations on to what they *think*, *know*, or *remember*, among other insight processes. Future work might explore how to co-opt this verbalization of insight to support deeper metacognitive practices in the presence of contextual factors.

Third, our exploration of pronouns beyond *I* indicated that the decrease in *I* pronouns was accompanied by an increase in third-person singular *he/she/it* pronouns. This suggests that the introduction of the contextual factor may be acting on clinical reasoning in part by distracting the participant away from her own reasoning actions and toward the actions of others (patient and, in the video condition, doctor). This finding further explains earlier work that found frequent mentions of contextual factors in think alouds[24,25]: the shift in focus to a contextual factor that is patient related entails a shift in focus to the patient rather than the diagnostic process *about* the patient.

Our study has several important limitations. First, think alouds are not a direct measure of cognition. Instead, they are an assessment method for understanding what individuals think based on what they say. Nonetheless, think alouds provide a useful way to explore clinical reasoning and linguistic markers.[3,23,44,48] Second, LIWC is not sensitive to context, and so it sometimes miscodes certain words (e.g., "*stress* test" as affective). This linguistic "noise," however, appears to be present across both conditions, thereby allowing LIWC to detect meaningful differences. Nonetheless, future work could benefit from refinements in the linguistic software. Finally, we examined think alouds across video and live scenario modalities. While this offers the power to discern differences, these difference between face to face and video interaction may impact clinical reasoning process and arguably are different contexts.

Our research has important implications for practice. As observed in the present study, context affects the clinical reasoning process, as predicted by situated cognition theory. Taken together with the research on errors in reasoning outcomes,[2–4] these findings argue for education around these contextual factors, perhaps through training in metacognition and awareness.[49] Moreover, the richness of these process-based measures of clinical reasoning lend themselves to

a more nuanced conceptualization of “performance.” These linguistic measures could be added to the growing assessment toolbox in medical education to improve early education and remediation of struggling learners. As voice recognition technology improves, automating transcription of learner reflections, LIWC could eventually be used as a formative assessment tool to alert instructors to when learners are being distracted by contextual factors and need support.

From a research perspective, these findings support the value of empirical work using situated cognition to explore context; reasoning differs in the presence of inhibiting contextual factors. Furthermore, while scholars are beginning to examine the cognitive processes and emotions inherent in physicians’ clinical reasoning,[9,11,20,21,50] to our knowledge, agency has not yet been addressed. Future research could investigate, for example, whether experiences of agency shift between clinic and inpatient contexts and, if so, whether this affects clinical reasoning.

Finally, these findings demonstrate the value of linguistic analysis generally and LIWC in particular. Such tools could be applied beyond the application of exploring context specificity, examining, for instance, errors present in electronic health records, assessment of diagnostic competencies, or patient-doctor communications. If we listen carefully to what physicians say about and during the diagnostic process, we may be able to better support them across shifting and even confusing contexts.

Acknowledgements: The authors would like to thank Megan Ohmer for her help in data collection and transcription.

Author Contributions:

All authors collaborated together on the research design and data collection. Konopasky ran the analyses and wrote up the results section. Konopasky and the remaining authors co-wrote the remainder of the paper. All authors offered substantive revisions and approve of this final version of the paper.

Research Funding: This study was supported by a grant from the JPC - 1, CDMRP - Congressionally Directed Medical Research Program (#NH83382416)

Competing Interest: The funding organization played no role in the study design; in the collection, analysis, and interpretation of data; in the writing of the report; or in the decision to submit the report for publication.

Disclaimers: The views expressed in this paper are those of the authors and do not necessarily reflect the official position or policy of the US Government, Department of Defense, Department of the Navy, or the Uniformed Services University of the Health Sciences.

References

1. Donaldson MS, Corrigan JM, Kohn LT. *To Err Is Human: Building a Safer Health System*. Vol 6. National Academies Press; 2000.
2. Singh H, Graber ML. Improving Diagnosis in Health Care--The Next Imperative for Patient Safety. *N Engl J Med*. 2015;373(26):2493.
3. Durning SJ, Artino AR, Boulet JR, Dorrance K, van der Vleuten C, Schuwirth L. The impact of selected contextual factors on experts' clinical reasoning performance (does context impact clinical reasoning performance in experts?). *Adv Heal Sci Educ*. 2012;17(1):65-79.
4. Eva KW. What every teacher needs to know about clinical reasoning. *Med Educ*. 2005;39(1):98-106.
5. Brown JS, Collins A, Duguid P. Situated cognition and the culture of learning. *Educ Res*. 1989;18(1):32-42.
6. King A. Structuring peer interaction to promote high-level cognitive processing. *Theory Pract*. 2002;41(1):33-39.
7. Kaptelinin V, Nardi BA. *Acting with Technology: Activity Theory and Interaction Design*. MIT press; 2006.
8. Konopasky AW, Sheridan KM. Towards a Diagnostic Toolkit for the Language of Agency. *Mind, Cult Act*. 2016;23(2). doi:10.1080/10749039.2015.1128952
9. Kozlowski D, Hutchinson M, Hurley J, Rowley J, Sutherland J. The role of emotion in clinical decision making: an integrative literature review. *BMC Med Educ*. 2017;17(1):255.
10. Stolper E, Van de Wiel M, Van Royen P, Van Bokhoven M, Van der Weijden T, Dinant

GJ. Gut feelings as a third track in general practitioners' diagnostic reasoning. *J Gen Intern Med*. 2011;26(2):197-203.

11. Resnick ML. The effect of affect: Decision making in the emotional context of health care. In: *2012 Symposium on Human Factors and Ergonomics in Health Care.* ; 2012:39-44.
12. Tausczik YR, Pennebaker JW. The psychological meaning of words: LIWC and computerized text analysis methods. *J Lang Soc Psychol*. 2010;29(1):24-54.
13. Pennebaker JW. The Secret Life of Pronouns: What Our Words Say About Us (NY: Bloomsbury). 2011.
14. Pennebaker JW, Graybeal A. Patterns of natural language use: Disclosure, personality, and social integration. *Curr Dir Psychol Sci*. 2001;10(3):90-93.
15. Konopasky A, Ramani D, Ohmer M, Durning SJ, Artino Jr AR, Battista A. Why health professions education needs functional linguistics: the power of 'stealth words.' *Med Educ*. 2019.
16. Bandura A. Toward a psychology of human agency. *Perspect Psychol Sci*. 2006;1(2):164-180.
17. Weiner B. An attributional theory of achievement motivation and emotion. *Psychol Rev*. 1985;92(4):548.
18. Simpkin AL, Murphy Z, Armstrong KA. A randomized experimental study to assess the effect of language on medical students' anxiety due to uncertainty. *Diagnosis*. 2019.
19. Damasio AR. *Descartes' Error*. Random House; 2006.
20. Heyhoe J, Birks Y, Harrison R, O'Hara JK, Cracknell A, Lawton R. The role of emotion in patient safety: are we brave enough to scratch beneath the surface? *J R Soc Med*.

2016;109(2):52-58.

21. Surry LT, Torre D, Trowbridge RL, Durning SJ. A mixed-methods exploration of cognitive dispositions to respond and clinical reasoning errors with multiple choice questions. *BMC Med Educ*. 2018;18(1):277.
22. Mamede S, Van Gog T, Schuit SCE, et al. Why patients' disruptive behaviours impair diagnostic reasoning: a randomised experiment. *BMJ Qual Saf*. 2017;26(1):13-18.
23. Estrada CA, Isen AM, Young MJ. Positive affect facilitates integration of information and decreases anchoring in reasoning among physicians. *Organ Behav Hum Decis Process*. 1997;72(1):117-135.
24. McBee E, Ratcliffe T, Picho K, et al. Consequences of contextual factors on clinical reasoning in resident physicians. *Adv Heal Sci Educ*. 2015;20(5):1225-1236.
25. Durning S, Artino Jr AR, Pangaro L, van der Vleuten CPM, Schuwirth L. Context and clinical reasoning: understanding the perspective of the expert's voice. *Med Educ*. 2011;45(9):927-938.
26. Kahn JH, Tobin RM, Massey AE, Anderson JA. Measuring emotional expression with the Linguistic Inquiry and Word Count. *Am J Psychol*. 2007:263-286.
27. Ilgen JS, Eva KW, Regehr G. What's in a Label? Is Diagnosis the Start or the End of Clinical Reasoning? *J Gen Intern Med*. 2016;31(4):435-437. doi:10.1007/s11606-016-3592-7
28. Konopasky AW, Ramani D, Ohmer M, et al. It Totally Possibly Could Be: How A Group Of Military Physicians Reflect On Their Clinical Reasoning In The Presence Of Contextual Factors. *Mil Med*.
29. McBee E, Ratcliffe T, Picho K, et al. Contextual factors and clinical reasoning:

differences in diagnostic and therapeutic reasoning in board certified versus resident physicians. *BMC Med Educ*. 2017;17(1):211.

30. Cleary TJ, Durning S, Gruppen L, Hemmer P, Artino A. Self-regulated learning in medical education. *Oxford Textb Med Educ*. 2013:465-477.
31. Patel V, Arocha J, Zhang J. Thinking and reasoning in medicine. In: Holyoak K, Morrison R, eds. *The Cambridge Handbook of Thinking and Reasoning*. Cambridge University Press; 2005:727-750.
32. ten Cate O, Durning SJ. Understanding Clinical Reasoning from Multiple Perspectives: A Conceptual and Theoretical Overview. 2018:35-46. doi:10.1007/978-3-319-64828-6_3
33. Ratcliffe TA, McBee E, Schuwirth L, et al. Exploring Implications of Context Specificity and Cognitive Load in Residents. *MedEdPublish*. 2017;6.
34. Paas F, Tuovinen JE, Tabbers H, Van Gerven PWM. Cognitive load measurement as a means to advance cognitive load theory. *Educ Psychol*. 2003;38(1):63-71.
35. Fraser K, Ma I, Teteris E, Baxter H, Wright B, McLaughlin K. Emotion, cognitive load and learning outcomes during simulation training. *Med Educ*. 2012;46(11):1055-1062.
36. Khawaja MA, Chen F, Marcus N. Measuring cognitive load using linguistic features: implications for usability evaluation and adaptive interaction design. *Int J Hum Comput Interact*. 2014;30(5):343-368.
37. Arnold J, Clarke DJ. What is ‘agency’? Perspectives in science education research. *Int J Sci Educ*. 2014;36(5):735-754.
38. Howe A, Barrett A, Leinster S. How medical students demonstrate their professionalism when reflecting on experience. *Med Educ*. 2009;43(10):942-951.
39. Battista A, Konopasky A, Ramani D, et al. Clinical reasoning in the primary care setting:

Two scenario-based simulations for residents and attendings. *MedEdPORTAL*. 2018;14(10773).

40. Durning SJ, Artino A, Boulet J, et al. The feasibility, reliability, and validity of a post-encounter form for evaluating clinical reasoning. *Med Teach*. 2012;34(1):30-37. doi:10.3109/0142159X.2011.590557

41. Ericsson KA, Simon HA. How to study thinking in everyday life: Contrasting think-aloud protocols with descriptions and explanations of thinking. *Mind, Cult Act*. 1998;5(3):178-186.

42. Durning SJ, Artino AR, Beckman TJ, et al. Does the think-aloud protocol reflect thinking? Exploring functional neuroimaging differences with thinking (answering multiple choice questions) versus thinking aloud. *Med Teach*. 2013;35(9):720-726. doi:10.3109/0142159X.2013.801938

43. Fox MC, Ericsson KA, Best R. Do procedures for verbal reporting of thinking have to be reactive? A meta-analysis and recommendations for best reporting methods. *Psychol Bull*. 2011;137(2):316.

44. Burbach B, Barnason S, Thompson SA. Using "think aloud" to capture clinical reasoning during patient simulation. *Int J Nurs Educ Scholarsh*. 2015;12(1):1-7.

45. Durning SJ, Graner J, Artino Jr AR, et al. Using functional neuroimaging combined with a think-aloud protocol to explore clinical reasoning expertise in internal medicine. *Mil Med*. 2012;177(suppl_9):72-78.

46. Funkesson KH, Anbäcken E-M, Ek A-C. Nurses' reasoning process during care planning taking pressure ulcer prevention as an example. A think-aloud study. *Int J Nurs Stud*. 2007;44(7):1109-1119.

47. Shaw BR, Hawkins R, McTavish F, Pingree S, Gustafson DH. Effects of insightful disclosure within computer mediated support groups on women with breast cancer. *Health Commun*. 2006;19(2):133-142.

48. Oliver EJ, Markland D, Hardy J, Petherick CM. The effects of autonomy-supportive versus controlling environments on self-talk. *Motiv Emot*. 2008;32(3):200-212.

49. Trowbridge RL, Dhaliwal G, Cosby KS. Educational agenda for diagnostic error reduction. *BMJ Qual Saf*. 2013;22(Suppl 2):ii28-ii32.

50. McBee E, Ratcliffe T, Goldszmidt M, et al. Clinical reasoning tasks and resident physicians: what do they reason about? *Acad Med*. 2016;91(7):1022-1028.

Effects of Live and Video Simulation on Clinical Reasoning Performance and Reflection

Clinical reasoning—the integration of clinical information, medical knowledge, and contextual factors to make patient care decisions[1–4]—is complex and has high-stakes implications for patients and society as a whole. In fact, diagnostic error is a major problem in healthcare accounting for approximately 10% of patient deaths as well as other adverse hospital events.[5–7] Although the research literature targeting clinical reasoning *outcomes* (i.e., diagnoses) continues to grow, we still know relatively little about the experiences and reactions of medical professionals during the clinical reasoning *process*, particularly regarding how such experiences relate to different approaches in simulation training environments (e.g., video case scenarios, scenario-based simulations employing simulated participants). To address this gap in the literature, we directly compared the effects of video case (i.e., pre-recorded video presentation of a doctor and patient encounter[8]) and live scenario-based simulation (i.e., simulations guided by a narrative set in a simulation room or clinical setting that rely on props and diverse actors to mimic a clinical situation[9]) on the clinical reasoning performance and task-specific reflective judgments of physicians.

Clinical Reasoning is Complex and Situated

Although clinical reasoning is often conceptualized as an end *product*, Ilgen, Eva, and Regehr argue that it is also a complex, dynamic, and often uncertain *process* of meaning making.[10] Goldszmidt and colleagues have begun to examine this process, identifying 26 distinct tasks that physicians may cycle through iteratively as they reason through a case (e.g., assessing priorities, identifying risk factors).[11–13] Ilgen and colleagues argue that the skillful deployment and completion of these tasks shift according to the case and the context, painting a complex and situation-specific (situated) picture of clinical reasoning.

Beyond the complexity of the clinical reasoning task itself, there is a developing literature on *contextual factors*—common features of clinical practice (e.g., patient frustration, interruption of clinician, physical space) outside of the case content; that is, factors that typically are not used to establish the correct diagnosis.[14–18] However, situated cognition theory and recent research suggests that when these contextual factors are present, physicians often think about and react to different aspects of a case, which can ultimately alter the quality or accuracy of their diagnostic and management reasoning.[14,18,19] For instance, McBee and colleagues showed that different types of patient contextual factors (i.e., limited patient language proficiency, patient challenging physician credentials) elicit negative emotional reactions in resident physicians which undermined their clinical reasoning.[20] Similarly, there is ample evidence to suggest that clinical reasoning is a situated endeavor governed by the facts needed to establish the diagnosis, as well as clinician factors (e.g., sleepiness, physician affective state), patient factors (e.g., patient expression of affect, diagnostic suggestion), and environmental factors (e.g., interruptions, electronic health record functionality, appointment length; see Figure 1).

Researchers have also examined the influence common *simulation training* approaches used to teach and evaluate clinical reasoning, such as live scenarios (i.e., scenario-based simulations), video (i.e., video cases), or virtual (i.e., computer-based) simulations, .[8,21–26] Research suggests that while all of these approaches ask participants to solve clinical problems, there are distinct differences in their characteristics, and in some instances, how these characteristics influence participant engagement behaviors, activities, and emotions. For example, live scenario-based simulations require participants to *simultaneously* engage in a wide range of complex clinically relevant activities (e.g., decision making, social interactions, structured interventions such as a focused assessment) while also determining the sequence of

these activities affording participants a great deal of autonomy and agency.[27] Similarly, virtual simulations also afford participants a central role in exercising what, when, and how they engage, but the simulation is typically encountered through a computer screen.[9] Conversely, video cases tend to less cognitively taxing, as participants simply view a pre-recorded encounter of another clinician interacting with a patient. Videos also typically have a fixed delivery time and present pre-determined content (e.g., interview, physician exam maneuvers, lab results) in a fixed order.[8] 2012). Based on these differences between live scenarios and video cases, it seems reasonable to speculate that participants in the live scenarios would experience greater cognitive load. From a cognitive load theoretical perspective, humans have a limited cognitive architecture and can typically only hold or process a limited amount of information in our short-term (or working) memory, classically defined as 7+2 (and more recently revised to 4+2).[28] Exceeding one's working memory capacity can impede learning and performance and lead to errors in clinical reasoning.

Broadly speaking, the literature is mixed regarding the superiority of any instructional approach across all contexts, outcomes, or situations. For example, while Durning and colleagues found no differences in clinical reasoning performance across standardized patient case, video case, and paper case formats,[23] LaRochelle and colleagues observed that standardized patient case and video case were superior to paper case formats, but only for certain subject areas.[8] Another important gap in the literature pertains to the cognitive experiences and underlying regulatory processes of medical professionals during simulation case experiences. Over the past decade, there has been increased interest in assessing the self-regulated learning (SRL) processes of medical professionals, such as planning, monitoring, and goal-setting as they engage in clinical reasoning.[29–35] As an assessment methodology, SRL microanalysis, which entails a

contextualized, structured interview approach, has emerged as a viable way to target these types of regulatory processes within the context of clinical reasoning and patient encounters.[29,30,32,36]

Purposes

The primary purpose of the current study was to experimentally examine the effects of live scenario-based simulations vs. video case approaches on physicians' reflective judgments (i.e., perceived challenges, adaptive inferences), perceptions of mental effort (i.e., cognitive load), and clinical reasoning performance. Microanalytic questions were used to assess two SRL processes: *perceived challenges* (i.e., perceptions of the primary difficulties experienced during a given activity) and *adaptive inferences* (i.e., conclusions made following task performance regarding how best to improve performance). A few hypotheses were generated for this study. Given the inherent differences in characteristics between live scenario-based simulations and video cases, we expected to observe some group differences across SRL processes. However, because research is scarce regarding process-related differences in this context, we did not speculate on the specific nature of these distinctions. Further, given the distinctions in characteristics between the two simulation approaches noted previously, we hypothesized that physicians in the live scenario would express higher perceptions of load than those from the video case. Finally, because participants in the live scenario would likely experience greater autonomy and agency for administering questions and examining the patient (as opposed to viewing an encounter and providing diagnostic and management decisions as with a video case),[27,29,30,33,34,36] we hypothesized that physicians from the live condition would perform at a higher level.

Method

Sample

This study was part of a larger clinical reasoning study conducted at three different military facilities across the U.S with 38 military physicians. The three military facilities are educational sites for the Uniformed Services University of the Health Sciences (USU) and represent regional tertiary referral centers of similar size for the military population. Physicians within the Military Health System frequently rotate between these and other hospitals.

Measures

Microanalytic questions. Participants were administered two microanalytic questions following completion of a post-encounter form (PEF): (a) perceived challenges, and (b) adaptive inferences. These questions were similar to those used in prior research[30] except for minor modifications to phrasing to best reflect current study objectives. As recommended by Cleary (2011), the microanalytic questions adhered to a free-response format.[37] Two individuals independently coded the responses from all 38 participants using a previously established coding scheme.[38,39] The raters discussed all instances of disagreement and used the lead author to make final determinations.

Perceived challenge. Consistent with microanalysis methodology, a single item was used to examine the perceptions of physicians regarding challenges encountered when completing the PEF and identifying the leading diagnosis ("*What was the most difficult thing for you when attempting to come up with the leading diagnosis*?"). The participants' responses were coded into one of the following five categories: (a) analysis of data, (b) knowledge/skill, (c) lack of case information, (d) no challenge, and (e) other.[30] (see Appendix for definitions and example

responses of each coding category). The inter-rater reliability for this measure was robust as indicated by a percent agreement of 98.2%.

Adaptive inferences. A single item measure was also used to assess the conclusions that the participants made regarding potential areas to adapt or improve upon when engaged in a similar professional activity in the future ("*Is there anything you would do differently when figuring out the leading diagnosis if you watched the video/participated in the scenario again*?") The coding scheme consisted of four broad categories: (a) general clinical tasks (i.e., history, testing, physical exam), (b) specific clinical reasoning sub-process (e.g., identifying symptoms, prioritizing symptoms, integration etc.), (c) none (i.e., no change was needed), and (d) other (see Appendix for category definitions and example responses). The inter-rater reliability for this measure was high (94.8%).

Perceived mental effort (cognitive load). Participants were asked to rate the level of mental effort used during the simulation experience. The participants were administered the prompt, "*Select your invested mental effort as you worked through the post-encounter form*", and asked to rate their effort using a 10-point Likert scale ranging from 1 (very low mental effort) to 10 (very high mental effort). This single item-measure of cognitive load has been used in prior studies and has been shown to reliably differentiate groups and to correlate with task difficulty and physiologic measures of cognitive load.[16,40–43]

Post-encounter form (PEF). A PEF developed in prior research was used as the basis for evaluating the quality of participants' clinical reasoning.[38,39] It consisted of seven open-ended scored sections (i.e., history questions, exam actions, problem list, differential diagnosis, leading diagnosis, supporting evidence, and management plan). We used a scoring instrument developed in prior research, (where inter-rater reliability was between kappa = .82 and kappa = .93 for all

sections).[38,39] An investigator matched free-text responses to the scoring sheet, which stipulated a score of correct (2 points), partially correct (1 point), or incorrect (0 points) for every potential response. These were all reviewed for accuracy by SJD, an internist, and, in the case of any novel responses, another internist reviewed them together with SJD and they came to consensus. These scores were then converted to percentage by dividing total number of points received by total possible score (e.g., if a participant gave two pieces of supporting evidence, they would have a total *possible* score of 4). For the purposes of this study, we calculated a single PEF score by averaging all the sections. Higher scores reflect stronger clinical reasoning performance. The coefficient alpha observed in this study for the PEF measure was .71.

Experimental design, procedures, and analysis

The current study was part of a broader investigation examining the effects of contextual factors across diagnosis type and simulation approaches.[15,44,45] Data from two conditions were used. The conditions were identical in terms of presenting condition (i.e., angina) and presence of contextual factors, but varied in simulation approach (live scenario vs. video case). Both conditions began with an informed consent process followed by a brief pre-study questionnaire. Participants in the video condition watched a video of the patient encounter that ranged in length from 3.47 to 4.23 minutes. They subsequently completed the PEF followed by the microanalytic and mental effort questions. The live scenario procedures were identical except for taking part in live scenarios with trained simulated patients. These scenarios ranged in length between 11 and 17 minutes (see Figure 2).

Inferential and descriptive statistics were used to address the research questions. Independent *t*-tests assessed group differences in clinical reasoning performance (i.e., PEF score) and perceived mental effort, while chi-square analyses examined group differences in self-

reflection processes (i.e., perceived challenges, adaptive inferences). Given the modest sample size used in this study, the likelihood ratio chi-square was used.[46] An a priori selected *p*-value of .05 was used for all inferential analyses. Finally, descriptive statistics (frequency counts, percentages) were used to examine the range and nature of responses across both microanalytic questions.

Results

The 38 participants were from different specialties (i.e., internal medicine (68.4%), family medicine (13.2%), and surgery (18.4%)) with varying level of expertise (i.e., intern (42.1%), resident (18.4%), and attending (39.5%)). The majority of the participants were male (65.8%), with an average age of approximately 36 years.

Video vs. Live Scenario Group Differences

Independent t-test revealed a statistically significant group difference for clinical reasoning performance (t (36) = 7.22, $p < .05$, $d = .42$) but not for perceived mental effort (t (36) = 1.31, p =.667) Thus, individuals from the live scenario condition outperformed those from the video condition even though they perceived their level of effort to be comparable to those from the video condition. The effect size for performance is considered medium.[47]

Chi-square analyses were conducted to investigate group differences across perceived challenges and adaptive inferences. Regarding perceived challenges, statistically significant group differences emerged for two coding categories: *analysis of data* (χ^2 (1) = 7.16, $p < .05$, ϕ = 0.43) and *lack of case information* (χ^2 (1) = 5.15, $p < .05$, $\phi = 0.36$). The effect sizes for these observed differences are considered medium.[47] Thus, physicians from the live condition identified *analyzing case information* as the primary challenge encountered whereas those in the

video condition appeared to be most concerned with *lack of case information* as a barrier to arriving at the correct diagnosis. In terms of descriptive analysis of aggregated group data across coding categories, difficulty with data analysis was clearly the most frequently endorsed challenge (n = 22, 57.9%), whereas the frequency of knowledge or abilities (n = 1; 2.6%) or no challenge (n = 0; 0.0%) was negligible.

Regarding adaptive inferences, no statistically significant group differences emerged. Subsequent descriptive analysis revealed that approximately 25% of the physicians emphasized the need to perform additional or more complete *clinical tasks*, such as patient history, tests or labs, and physical exam, but only 3% (n = 2) focused on the need to engage in specific clinical reasoning sub-processes (e.g., identifying and prioritization of symptoms, integration of symptoms; see Table 2). Of particular interest, however, was that 50% (n = 19) of the physicians did not believe they needed to do anything differently to improve their performance (i.e., adaptive inference question). To further evaluate these “non-adaptation” responses, we used expert consensus for performance-based information (i.e., scores from leading diagnosis, supporting evidence, and management components of the PEF) to identify physicians who performed at an acceptable or subpar level. An *acceptable* performance level was defined as a score of at least 50% across all three PEF components, while a *subpar* designation involved a score of less than 50% on any of these components. Approximately 42% (n = 8) of the physicians who provided a non-adaptation response exhibited subpar performance; suggesting that a fairly large percentage of physicians would not change or do anything differently for the case even though they underperformed.

Discussion

This study is important because it sheds light on the differential effects of simulation training approaches across performance and reflective thinking measures during clinical reasoning. It also expands on previous microanalysis research in medical education by focusing on more experienced physicians, as well as on the nature and quality of reflection-phase processes during clinical reasoning.[30,32]

Simulation Group Differences

One of the most notable findings from this study was that physicians who engaged patients in the live condition outperformed and identified different types of challenges than those who watched a video of a patient exhibiting the same condition (i.e., angina). Although data from this study cannot be used to make broad claims about the superiority of live simulation training experiences, to our knowledge, this study is one of the first to examine the relative benefits of different simulation approaches in terms of physician perspectives and underlying processes. Specifically, this study directly assessed physician perceptions of reflective judgments regarding the need to adapt or change their own behaviors during clinical reasoning.

Several points regarding the interplay of cognitive and SRL processes during clinical reasoning should be addressed. First, the fact that physicians reported different types of challenges across conditions suggests that simulation approaches may inherently direct or focus participant attention on different aspects of the case. Of particular interest was a fairly high percentage of video case participants (n = 8; 42%), but not live scenario (n = 2; 10%), who identified *lack of case information* as their primary challenge. This finding was provocative because the group experts who created the video noted that it contained all of the relevant information needed to identify the correct diagnosis; thereby suggesting that many video

participants were either unaware of or not knowledgeable about the key pieces of information related to the condition. This premise is supported by research showing that physicians often miss key information when viewing videos of patient encounters with contextual factors.[16,20] Conversely, the physicians in the live scenario condition provided more "adaptive" perceived challenge responses in that they were thinking primarily about the need to engage in more effective data analysis processes (e.g., integrating symptoms etc.). Although tentative at this point, our results collectively suggest that different simulation approaches differentially influence specific focus or thought processes of physicians during a clinical encounter

A second point related to SRL processes was that, regardless of condition, 50% of the total sample (n = 19) revealed that they would not change or do anything differently with the case, with approximately 42% (n = 8) of that subgroup exhibiting subpar performance. Thus, approximately 21% of the total sample indicated that they would not change or do anything differently with the case even though they did not perform well – indicating mis-calibration or another type of self-assessment problem. In other words, in this study, a sizeable percentage of physicians did not appear to accurately assess their level of performance, or the actions needed to optimize performance. This finding is consistent with the medical education literature showing that medical professionals often exhibit inconsistencies in their own metacognitive judgments of performance and show differences in the accuracy of these judgments across different part of clinical reasoning activities.[30,48,49]

Another key issue involved the nature of physicians' adaptive inferences was lack of focus on specific clinical reasoning sub-processes (i.e., identifying symptoms, integration, comparing and contrasting diagnoses). For example, approximately 25% of the participants provided very broad responses regarding the need to conduct more tests or to do a more thorough

patient exam and physical exam, while only 5% (n = 2) of the physicians mentioned specific clinical reasoning sub-processes as an important area to focus on. When also considering that 50% of the physicians reported that changes or modifications were *not needed*, these results are consistent with previous SRL microanalytic research showing that medical students do not consistently think about focus on such processes at the outset of a patient encounter, and often abandon process-oriented ways of thinking when challenges are encountered.[29,32]

Finally, we hypothesized that due to the greater demands and expectations for self-directedness, physicians from the live scenario condition would report experiencing greater cognitive load than those from the video case. This hypothesis was not supported. While it is true that individuals during live or virtual simulation encounters bear the burden of directly managing the clinical encounter (and thus experiencing greater cognitive load), it seems likely that physicians in the video condition experienced similar mental effort demands, albeit for a different reasons. It possible that the perceived lack of case information during the video condition led the physicians to expend a level of mental effort comparable to the live scenario. .[50] Examining differences in the factors influencing cognitive load perceptions across simulation approaches is a potentially fruitful line of research may represent an area to improve clinical reasoning and reduce error.

Limitations

Although these results are informative, there are a few limitations that warrant attention. First, the modest sample size prevented us from including moderator variables in the analyses, such as physician experience level, presence or absence of contextual factors, or type of condition. Thus, the external validity of this study is limited as it only focused on type of diagnosis (unstable angina) and included only more experienced physicians. Another limitation

was the use of a single item, self-report measure of cognitive load. Although this measure has an empirical foundation,[38,39] it is nonetheless highly simplistic and based purely on the participants' self-perceptions. Future work in this area should consider the use of more complete self-report measures and/or physiological correlates of cognitive load.[51] Finally, in this study we only examined physician performance and reflection using a single case. Due the challenges related to case specificity, investigators should consider the use of multiple cases in future research.

Conclusions

The current study sheds light on the differential effects of simulation approach and the benefits of using context-specific assessment tools to uncover underlying cognitive judgments and reactions of medical professionals during clinical tasks. Although more research is clearly need, revelations regarding cognitive and regulatory processes can prove to be quite valuable to educators, researchers, and others interested in providing personalized, process-oriented feedback to trainees and even practicing physicians.

References

1. Croskerry P. A universal model of diagnostic reasoning. *Acad Med*. 2009;84(8):1022-1028.
2. Elstein AS, Shulman LS, Sprafka SA. Medical problem solving an analysis of clinical reasoning. 1978.
3. Heneghan C, Glasziou P, Thompson M, et al. Diagnostic strategies used in primary care. *Bmj*. 2009;338(apr20_1):b946-b946.

4. Young M, Thomas A, Lubarsky S, et al. Drawing boundaries: the difficulty in defining clinical reasoning. *Acad Med*. 2018;93(7):990-995.

5. Singh H, Graber ML. Improving Diagnosis in Health Care--The Next Imperative for Patient Safety. *N Engl J Med*. 2015;373(26):2493.

6. National Academies of Sciences and Medicine E. *Improving Diagnosis in Health Care*. National Academies Press; 2016.

7. Trowbridge RL, Dhaliwal G, Cosby KS. Educational agenda for diagnostic error reduction. *BMJ Qual Saf*. 2013;22(Suppl 2):ii28-ii32.

8. LaRochelle JS, Durning SJ, Pangaro LN, Artino AR, van der Vleuten C, Schuwirth L. Impact of increased authenticity in instructional format on preclerkship students' performance: a two-year, prospective, randomized study. *Acad Med*. 2012;87(10):1341-1347.

9. Lopreiato JO. *Healthcare Simulation Dictionary*. Agency for Healthcare Research and Quality; 2016.

10. Ilgen JS, Eva KW, Regehr G. What's in a Label? Is Diagnosis the Start or the End of Clinical Reasoning? *J Gen Intern Med*. 2016;31(4):435-437. doi:10.1007/s11606-016-3592-7

11. McBee E, Ratcliffe T, Goldszmidt M, et al. Clinical reasoning tasks and resident physicians: what do they reason about? *Acad Med*. 2016;91(7):1022-1028.

12. Juma S, Goldszmidt M. What physicians reason about during admission case review. *Adv Heal Sci Educ*. 2017;22(3):691-711.

13. Goldszmidt M, Minda JP, Bordage G. Developing a unified list of physicians' reasoning tasks during clinical encounters. *Acad Med*. 2013;88(3):390-394.

14. Durning SJ, Artino AR, Boulet JR, Dorrance K, van der Vleuten C, Schuwirth L. The impact of selected contextual factors on experts' clinical reasoning performance (does context impact clinical reasoning performance in experts?). *Adv Heal Sci Educ*. 2012;17(1):65-79.

15. Konopasky AW, Ramani D, Ohmer M, et al. It Totally Possibly Could Be: How A Group Of Military Physicians Reflect On Their Clinical Reasoning In The Presence Of Contextual Factors. *Mil Med*.

16. Ratcliffe TA, McBee E, Schuwirth L, et al. Exploring Implications of Context Specificity and Cognitive Load in Residents. *MedEdPublish*. 2017;6.

17. Westbrook JI, Raban MZ, Walter SR, Douglas H. Task errors by emergency physicians are associated with interruptions, multitasking, fatigue and working memory capacity: a prospective, direct observation study. *BMJ Qual Saf*. 2018:bmjqs-2017.

18. Mercuri M, Sherbino J, Sedran RJ, Frank JR, Gafni A, Norman G. When guidelines don't guide: the effect of patient context on management decisions based on clinical practice guidelines. *Acad Med*. 2015;90(2):191-196.

19. Durning S, Artino Jr AR, Pangaro L, van der Vleuten CPM, Schuwirth L. Context and clinical reasoning: understanding the perspective of the expert's voice. *Med Educ*. 2011;45(9):927-938.

20. McBee E, Ratcliffe T, Picho K, et al. Consequences of contextual factors on clinical reasoning in resident physicians. *Adv Heal Sci Educ*. 2015;20(5):1225-1236.

21. Cook DA, Brydges R, Zendejas B, Hamstra SJ, Hatala R. Technology-enhanced simulation to assess health professionals: a systematic review of validity evidence, research methods, and reporting quality. *Acad Med*. 2013;88(6):872-883.

22. Bearman M. Is virtual the same as real? Medical students' experiences of a virtual patient. *Acad Med*. 2003;78(5):538-545.

23. Durning SJ, Dong T, Artino Jr AR, et al. Instructional authenticity and clinical reasoning in undergraduate medical education: a 2-year, prospective, randomized trial. *Mil Med*. 2012;177(suppl_9):38-43.

24. DeMaria Jr S, Bryson EO, Mooney TJ, et al. Adding emotional stressors to training in simulated cardiopulmonary arrest enhances participant performance. *Med Educ*. 2010;44(10):1006-1015.

25. Harwayne-Gidansky I, Bellis JM, McLaren SH, et al. Mannequin-based immersive simulation improves resident understanding of a clinical decision rule. *Simul Gaming*. 2017;48(5):657-669.

26. Schuelper N, Ludwig S, Anders S, Raupach T. The Impact of Medical Students' Individual Teaching Format Choice on the Learning Outcome Related to Clinical Reasoning. *JMIR Med Educ*. 2019;5(2):e13386.

27. Battista A. An activity theory perspective of how scenario-based simulations support learning: a descriptive analysis. *Adv Simul*. 2017;2(1):23.

28. Van Merriënboer JJG, Sweller J. Cognitive load theory in health professional education: design principles and strategies. *Med Educ*. 2010;44(1):85-93.

29. Artino Jr AR, Cleary TJ, Dong T, Hemmer PA, Durning SJ. Exploring clinical reasoning in novices: a self-regulated learning microanalytic assessment approach. *Med Educ*. 2014;48(3):280-291.

30. Cleary TJ, Konopasky A, La Rochelle JS, Neubauer BE, Durning SJ, Artino AR. First-year medical students' calibration bias and accuracy across clinical reasoning activities.

Adv Heal Sci Educ. 2019:1-15.

31. Callan GL, Cleary TJ. Multidimensional assessment of self-regulated learning with middle school math students. *Sch Psychol Q*. 2018;33(1):103.
32. Cleary TJ, Dong T, Artino AR. Examining shifts in medical students' microanalytic motivation beliefs and regulatory processes during a diagnostic reasoning task. *Adv Heal Sci Educ*. 2015;20(3):611-626.
33. Patel R, Sandars J, Carr S. Clinical diagnostic decision-making in real life contexts: a trans-theoretical approach for teaching: AMEE Guide No. 95. *Med Teach*. 2015;37(3):211-227.
34. Sandars J. The use of reflection in medical education: AMEE Guide No. 44. *Med Teach*. 2009;31(8):685-695.
35. Sandars J. Pause 2 Learn. *Med Educ*. 2010;44(11):1122-1123.
36. Sandars J, Cleary TJ. Self-regulation theory: applications to medical education: AMEE Guide No. 58. *Med Teach*. 2011;33(11):875-886.
37. Cleary TJ. Emergence of self-regulated learning microanalysis. *Handb self-regulation Learn Perform*. 2011;1:329-345.
38. Durning SJ, Artino A, Boulet J, et al. The feasibility, reliability, and validity of a post-encounter form for evaluating clinical reasoning. *Med Teach*. 2012;34(1):30-37. doi:10.3109/0142159X.2011.590557
39. McBee E, Ratcliffe T, Picho K, et al. Contextual factors and clinical reasoning: differences in diagnostic and therapeutic reasoning in board certified versus resident physicians. *BMC Med Educ*. 2017;17(1):211.
40. Paas F, Tuovinen JE, Tabbers H, Van Gerven PWM. Cognitive load measurement as a

means to advance cognitive load theory. *Educ Psychol*. 2003;38(1):63-71.

41. Paas FGWC, Van Merriënboer JJG, Adam JJ. Measurement of cognitive load in instructional research. *Percept Mot Skills*. 1994;79(1):419-430.
42. Szulewski A, Gegenfurtner A, Howes DW, Sivilotti MLA, van Merriënboer JJG. Measuring physician cognitive load: Validity evidence for a physiologic and a psychometric tool. *Adv Heal Sci Educ*. 2017;22(4):951-968.
43. Young JQ, Irby DM, Barilla-LaBarca M-L, ten Cate O, O'Sullivan PS. Measuring cognitive load: mixed results from a handover simulation for medical students. *Perspect Med Educ*. 2016;5(1):24-32.
44. Alexis Battista Divya Ramani, Megan Ohmer, Jeffrey Mikita, Anna Howle, Sarah Krajnik, Dario Torre, Steven J Durning AK. Clinical Reasoning in the Primary Care Setting: Two Scenario-Based Simulations for Residents and Attendings. *MedEdPORTAL Publ*.
45. Ohmer M, Durning SJ, Kucera W, et al. Clinical Reasoning in the Ward Setting: A Rapid Response Scenario for Residents and Attendings. *MedEdPORTAL*. 2019;15.
46. Meyers LS, Gamst G, Guarino AJ. *Applied Multivariate Research: Design and Interpretation*. Sage publications; 2016.
47. Cohen J. *Statistical Power Analysis for the Behavioral Sciences*. Routledge; 1988.
48. Eva KW, Regehr G. Exploring the divergence between self-assessment and self-monitoring. *Adv Heal Sci Educ*. 2011;16(3):311-329. doi:10.1007/s10459-010-9263-2
49. Norman GR, Shannon SI, Marrin ML. Learning in practice: The need for needs assessment in continuing medical education. *Br Med J*. 2004;328:999-1001. doi:10.1136/bmj.328.7446.999

50. Sweller J, Ayres PL, Kalyuga S, Chandler P. The expertise reversal effect. 2003.

51. Solhjoo S, Haigney MC, McBee E, et al. Heart Rate and Heart Rate Variability Correlate with Clinical Reasoning Performance and Self-Reported Measures of Cognitive Load. *Sci Rep*. 2019;9(1):1-9.

Table 1

Frequency and Percentage of Perceived Challenge Responses Across Instructional Group

Perceived Challenge	Video (*n* =19) *n* (%)	Live (*n* = 19) *n* (%)	Total (*n* = 38) *n* (%)
Knowledge and skill	1 (5.2%)	0 (0%)	1 (2.6%)
Analysis of data	7 (36.8%)	15 (78.9%)	22 (57.9%)
Lack of case information	8 (42.1%)	2 (10.5%)	10 (26.3%)
No/none	0 (0%)	0 (0%)	0 (0%)
Other	3 (15.7%)	3 (15.7%)	6 (15.8%)

Note. The sum percentage for the Live condition exceeded 100% given that one participant in this group provided two codeable responses.

Table 2

Frequency and Percentage of Adaptive Inference Responses Across Instructional Group

Adaptive Inference	Video (*n* =19) *n* (%)	Live (*n* = 19) *n* (%)	Total (*n* = 38) *n* (%)
Clinical tasks (history, tests, exam)	5 (26.3%)	4 (21.1%)	9 (23.7%)
Process (Total)	0 (0%)	2 (10.5%)	2 (5.2%)
- identify symptoms	0 (0%)	1 (5.3%)	1 (2.6%)
- history and demographics	0 (0%)	1 (5.3%)	1 (2.6%)
- clarifying/prioritizing symptoms	0 (0%)	0 (0%)	0 (0%)
- integration	0 (0%)	0 (0%)	0 (0%)
- comparing/contrasting diagnoses	0 (0%)	0 (0%)	0 (0%)
None	9 (47.4%)	10 (52.6%)	19 (50.0%)
Other	5 (26.3%)	3 (15.8%)	8 (21.1%)

Figure 1. Contextual factors in clinical reasoning

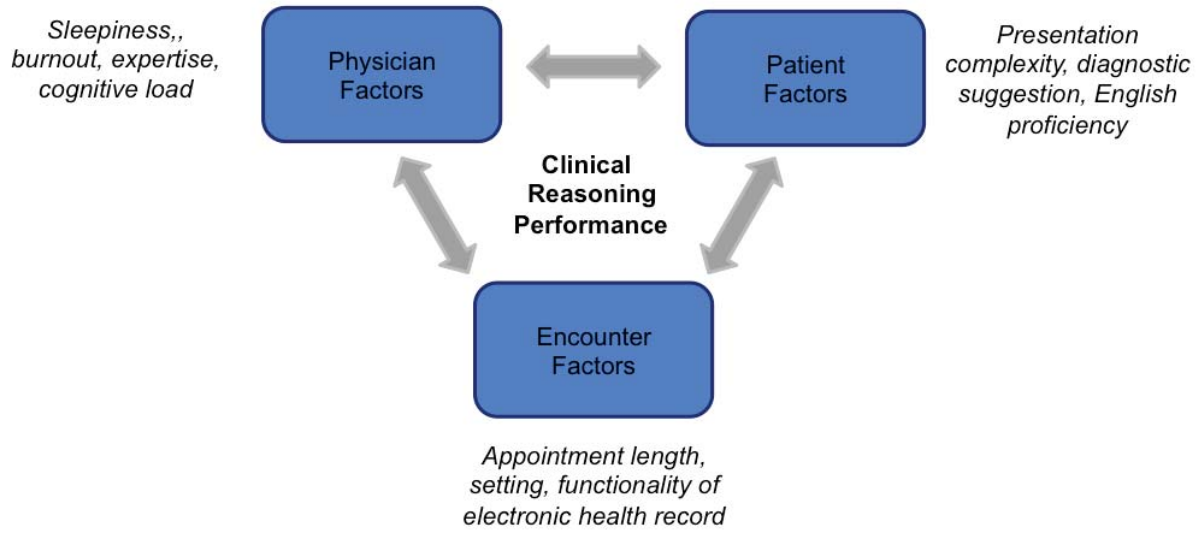


Figure 2. Workflow diagram of video-based and live scenario-based simulation study procedures.

Table of Contents
Supplemental Documents

20	Intervention Rationale
21	The Use of Think-Aloud Reflections to Examine Learners Experiences in Live and Video-Based Simulation Contexts: A Comparison Study [Presentation]
22	Examining the Influence of Simulation Context on Learners' Post-Simulation Reflections [Presentation]
23	Emergent Clinical Reasoning During ThinAk louds: How Physicians Reflect on their Own and Others' Practices in Live and Video Simulation [Presentation]
24	Case Specificity in Clinical Reasoning: A Qualitative Case Study of Conditional Reasoning Processes [Presentation]
25	An Introductory Workshop for Activity and Linguistic Analysis of Video in Healthcare Simulation
26	It Totally Possibly Could Be: How a Group of Military Physicians Reflect on Their Clinical Reasoning in the Presence of
27	Clinical Reasoning in the Medical/Surgical Ward Setting: A Rapid Response Scenario for Residents and Attendings
28	Clinical Reasoning in the Primary Care Setting: Two Scenario-Based Simulations for Residents and Attendings
29	Sequence Matters: Patterns in Task-Based Clinical Reasoning Short title: Sequence Matters
30	Examining the Patterns of Uncertainty Across Clinical Reasoning Tasks: Effects of Contextual Factors on Clinical Reasoning Performance
31	Understanding Context Specificity: The Effect of Contextual Factors on Clinical Reasoning
32	The Linguistic Effects of Context Specificity: Exploring Affect, Cognitive Processing, and Agency in Physicians' Think-Aloud Reflections

33	Heart Rate and Heart Rate Variability Correlate with Clinical Reasoning Performance and Self-Reported Measures of Cognitive Load
34	First-year medical students' calibration bias and accuracy across clinical reasoning activities: An initial investigation
35	Well-Being in a Cohort of Active Duty Military Physicians: An Assessment of Sleep Patterns, Burnout and Perceived Stress
36	The Linguistic Effects of Context Specificity: Exploring Affect, Cognitive Processes, and Agency in Physicians' Think-Aloud Reflections
37	Five Principles for Using Educational Theory: Strategies for Advancing Health Professions Education Research [Manuscript]
38	Scenario _ Trauma
39	Management Reasoning _ Jama Paper