Factors and Interactions Influencing Direct Observation: A Literature Review Guided by Activity Theory

By

Margaret M Swanberg, DO

Thesis submitted to the Faculty of the

Health Professions Education Graduate Program

Uniformed Services University of the Health Sciences

In partial fulfillment of the requirements for the degree of

Master's in Health Professions Education 2021

Distribution Statement

Distribution A: Public Release.

The views presented here are those of the author and are not to be construed as official or reflecting the views of the Uniformed Services University of the Health Sciences, the Department of Defense or the U.S. Government.

UNIFORMED SERVICES UNIVERSITY, SCHOOL OF MEDICINE GRADUATE PROGRAMS
Graduate Education Office (A 1045), 4301 Jones Bridge Road, Bethesda, MD 20814

May 20, 2021

APPROVAL SHEET

Title of Dissertation: **Factors and Interactions Influencing Direct Observation: A Literature Review Guided by Activity Theory**

Name of Candidate: **Margaret M. Swanberg**, Master of Science in **Health Professions Education**

02/23/2021

THESIS AND ABSTRACT APPROVED:

DATE:

[redacted] 02/23/2021

Michael Soh
CHPE, DEPARTMENT OF MEDICINE
Committee Chairperson

[redacted] 4March2021

Louis Pangaro
CHPE, DEPARTMENT OF MEDICINE
Thesis Advisor

[redacted] 3/4/2021

Lauren Maggio
CHPE, DEPARTMENT OF MEDICINE
Committee Member

[redacted] 03/03/2021

Dario Torre
CHPE, DEPARTMENT OF MEDICINE
Committee Member

Acknowledgements

I am particularly indebted to my thesis advisors, Dr. Lauren Maggio, Dr. Louis Pangaro, and Dr. Dario Torre, who provided me with countless hours of their time, their expertise, their mentorship, and their encouragement. Their dedication to this project has helped advance my personal and professional growth as an education researcher, a writer, and a clinician.

I am also grateful for the assistance of the Learning Resource Center, whose staff were a tremendous asset in conducting the literature reviews.

Lastly, I would like to acknowledge the help of LT Sarah Woodson, a future leader in education research. Her fresh perspectives and insightful comments were always thought provoking.

Dedication

I would like to dedicate this thesis to my two amazing children, Caitie and Sebastian. They are both brilliant and amazing young adults, and without their own personal and academic successes, I would not have been able to pursue this degree. They continue to remind me to never stop learning and to never completely grow up.

I would also like to dedicate this thesis to the numerous medical students and residents I have had the pleasure to teach. Their commitment to excellence inspires me to be a better educator so that I can continue to meet the needs of the current and future generation of military physicians.

Disclaimer

The views presented here are those of the author and are not to be construed as official or reflecting the views of the Uniformed Services University of the Health Sciences, the Department of Defense or the U.S. Government.

The author hereby certifies that the use of any copyrighted material in the thesis manuscript entitled:

Factors and interactions influencing direct observation: an Activity Theory analysis of the literature

is appropriately acknowledged and, beyond brief excerpts, is with the permission of the copyright owner.

SWANBERG.MARGARET.MARY.1035473030 Digitally signed by SWANBERG.MARGARET.MARY.1035473030
Date: 2021.05.20 13:32:01 -04'00'

Margaret M Swanberg
Health Professions, Education, Dept of Medicine
Uniformed Services University
Date: 05/17/2021

Abstract

Title: Factors and interactions influencing direct observation: A Literature Review Guided by Activity Theory

Purpose

Ensuring that future physicians are competent to practice medicine is necessary for high quality patient care and safety. The shift towards competency based education has placed renewed emphasis on direct observation in authentic patient care contexts. Despite this interest and multiple studies focused on improving direct observation, threats to the validity and reliability of measured outcomes remain underexplored and unresolved. I chose to re-examine the literature on direct observation in authentic contexts using Activity Theory as a guide to explore the numerous interactions and influences that may affect this complex activity.

Method:

I conducted a literature review of direct observation in authentic patient contexts by systematically searching databases PubMed, ERIC, Embase, and Web of Science. I extracted key data from the studies and applied Activity Theory to the findings to identify factors affecting these observations and the interactions between them.

Results:

19 articles were included in the analysis, 13 involved GME learners and 6 for UME. Of the 19 six were in the operating room setting and four were in the Emergency department. Using Activity Theory, I discovered that while numerous studies focus on rater and tool influences, very few study the impact of social elements. These are the rules that govern how the activity happens, the environment and members of the community involved in the activity and how completion of the activity is divided up among the members of the community. I also did not find a wealth of literature that examined the potential impact or influence of gender or ethnicity of the rater, learner, or patient.

Conclusions:

Viewing direct observation through the lens of Activity Theory enables educators to implement curricular changes to improve direct observation of assessment. Using Activity Theory as a guide, the multi-directional impact of the numerous elements involved in the performance of direct observation can be explored. Activity Theory also allows researchers to design studies to focus on the identified underexplored interactions and influences.

TABLE OF CONTENTS

LIST OF TABLES……viii

LIST OF FIGURES……..ix

CHAPTER 1: BACKGROUND…… .1

PURPOSE AND RESEARCH QUESTION ……..3

THEORETICAL FRAMEWORK……...4

CHAPTER 2: SUBMITTED MANUSCRIPT……..7

ABSTRACT…….7

BACKGROUND……..8

METHODS……11

RESULTS…….13

DISCUSSION……20

CHAPTER 3: DISCUSSION……..24

MILITARY RELEVANCE……27

CONCLUSION……..28

REFERENCES……29

LIST OF TABLES

Table 1: Summary characteristics of included articles……14

LIST OF FIGURES

Figure 1: Application of Activity Theory to COVID Quarantine……..5

Figure 2: Basic elements within an activity system ……11

Figure 3: Literature review flow chart ……14

Chapter 1: Introduction

Directly observing medical trainees performing essential physician skills, including interactions with patients, requesting consultation, running an interdisciplinary team or leading a family meeting has been used for decades in medical education as a method to assess learner abilities. However, due to the nature of direct observation, which is based on a human's perspective of the observed activity, there is potential for this type of assessment to be fraught with potential error and bias (1-3). Moreover, direct observation does not happen in isolation, but can be influenced by the environment (4,5), by prior experiences of the rater and learner (6), by members of the healthcare team and/or patient's family, and by the purpose of observation (7). In this light, direct observation can be considered a complex activity, subject to numerous influences beyond the learner and the rater. Therefore, to examine the varied influences on direct observation, I chose to apply the lens of Activity Theory (8,9) to explore published studies about direct observation in authentic clinical contexts. Activity Theory is a relatively new approach to analyzing the various elements involved in a complex activity system, such as occurs in direct observation, and the potential tensions between those elements. Through this lens, I hoped to characterize the literature, identify any gaps, and, if warranted, offer practical and evidence-informed suggestions for educators involved in direct observation.

Over the past several years there has been increasing focus on competency based medical education in both undergraduate and graduate medical education. The term competency based medical education (CBME) is subject to various definitions. In general CBME refers to medical education that is outcome based and geared towards observing and documenting a trainee's ability to perform the duties associated with their profession regardless of context. (10). Multiple frameworks have been proposed to assist medical educators in the assessment of medical trainees. (11) For example, the Accreditation Council for Graduate Medical Education, through the Milestones project (12) and the Association of American Medical Colleges' Entrustable Professional Activities (EPA) (13), both emphasize observable

skills and actions that a medical learner should perform as they advance towards greater independence. As medical students progress from pre-clerkship experiences towards advanced clerkship rotations, assessment tools shift focus from question-based exams towards direct observation, or competency-based examinations. These performance assessments are rooted in observations such as the ability to perform history taking, completion of a physical exam, and presentation of data to a resident or faculty member. Although knowledge is still assessed in clerkship students via National Board of Medical Examiners clinical subject exams or other multiple choice question exams, a significant component of whether or not the student passes the clerkship and advances to the next level is based on direct observation performance evaluations.

Direct observation of the student, intern, or resident when the attending is present during the entirety of the observed encounter is one of the most effective methods for ensuring that the learner possesses the knowledge, skills, and attitudes necessary to practice medicine in an authentic clinical context (2). This translates to the concept of *does* according to Miller's pyramid (14) and represents greater advancement along the trajectory from novice to expert and the move toward independent practice. However, despite the importance and effectiveness of this assessment approach, direct observation, like all assessments, presents challenges. For example, acknowledging that direct observation in authentic contexts is preferable in that it increases the content validity of the assessment (15), the use of real patient encounters for assessment can be challenging and can frequently be dependent on the randomness of patient complexity arriving in clinic that day, the time allocated for the encounter, patient behavior, and whether or not the patients are on time. Additionally, researchers have observed that in direct observation assessments much of the variability in learner performance can be attributed to the rater and not what knowledge, skills, and attitudes the trainee demonstrated. (2,16-20) Similarly, the impact of these factors has been noted in relation to learner characteristics.(21-23) Another issue is that the quality of the direct observation can be poor, even after preceptors have had a block of instruction on the use of the evaluation tool, face-to-face training on direct observation, and knowledge checks using

video assessment of learners; whereas in some cases training has been shown to improve observer performance. (24-26)

These numerous sources of variation; many of which are beyond the learner's control, can significantly impact the learner's confidence, performance, and even threaten the overall validity of the assessment outcome. Clerkship directors are tasked with making pass/fail decisions regarding a students' progress towards readiness for GME, which includes assessment of the student's ability to independently perform essential physician skills. Deans of medical schools are tasked with making decisions to award a medical degree to a student in part based on mastery of the skills deemed essential for a student to enter graduate medical education. Residency Program Directors in turn are tasked with graduating learners who have been deemed competent to independently practice their chosen specialty, in large part based on clinical observation of that learner's mastery of ACGME Milestones. Entrustment decisions about skills and decisions to allow a physician to practice independently without supervision require accurate, objective, and valid assessments of the physician's clinical skills. Direct observation can be applied to any of the physician skills deemed necessary by the EPA's and the Milestones, and can be a powerful tool to aid physicians tasked with making decisions regarding a learners' readiness to advance to the next academic level. When coupled with appropriate feedback direct observation can be a valuable instructional method for learning providing community members with insight into actions, behaviors, and knowledge. In the era of CBME, this method can be an integral part of summative assessments to determine if a learner has successfully mastered physician skills based on observing what the learner actually does in authentic clinical contexts. However, direct observation should play such a significant role when the outcome, whether it is for learning or of learning, is rooted in the individual knowledge, skills, and attitudes of the learner.

Purpose and Research Question

Several studies have examined factors related to rater cognition, rater training, and the assessment tools used to facilitate direct observation. To summarize these studies, knowledge syntheses have been

published exploring the literature on each of these subjects or their combination. For example, reviews have studied contextual factors such as the patient or case complexity as well as the environment in which the observation happens and time demands on the rater. (3,15,26) These knowledge syntheses generally examine the impact of these factors on the rater and rater assessment. However, to better understand what impacts direct observation assessment and why outside influences impact the outcome, the interaction of all the elements upon each other should be explored. Thus, I wanted to take a closer look at direct observation using a wider lens to investigate what has been published on these factors and to explore the interaction of other factors as they relate to this complex activity.

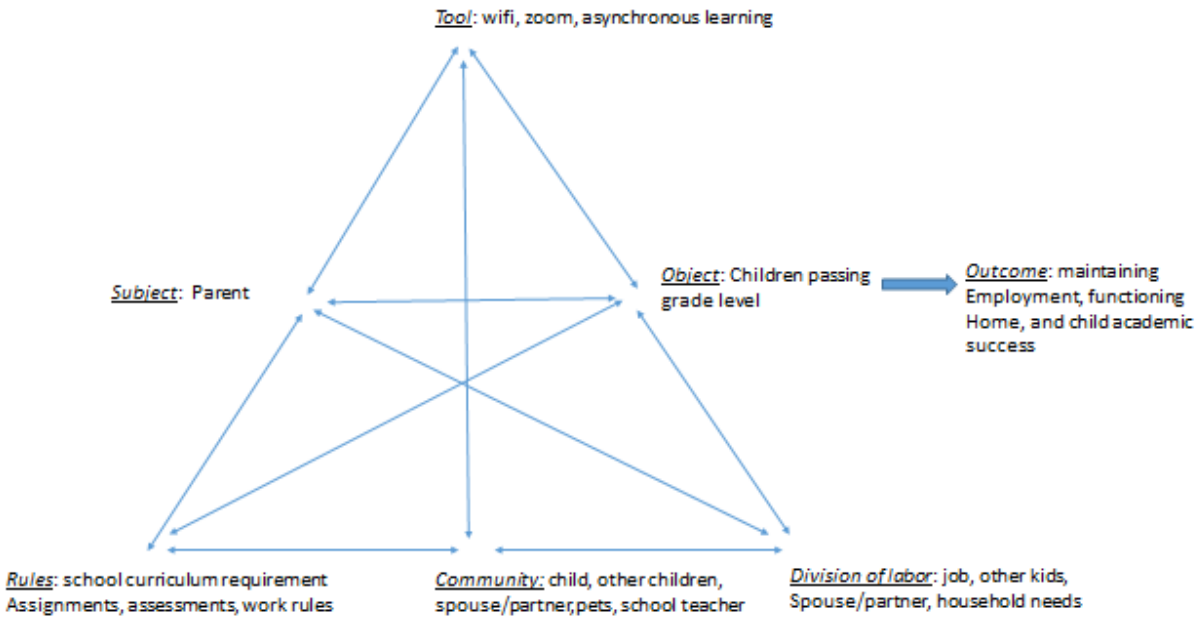
Theoretical framework

What is activity theory?

Activity Theory provides a framework to understand and analyze the complex activities of humans, the systems in which they work, and the interactions or tensions between multiple factors associated with the activity. (8,9) The key concept of Activity Theory is that there is a subject (the person doing the activity) who utilizes tools (either tangible or intangible) to accomplish the activity, to work on an object (usually described as the reason or purpose of the activity) in order to achieve an outcome, which is not necessarily the intended one. A central premise of AT is that knowledge and learning are mediated through the use of tools or artifacts. Additionally, the theory posits that the activity can be any interaction between the subject and the world around him/her. Another fundamental element of AT is that the subject and the object should both be transformed as a result of their interaction. The tools that help mediate the action can also be changed and developed over the course of the interactions between the subject and object of the activity. In the late 1990's, Engstrom expanded on AT and included elements related to the social context of the activity; the community, rules and division of labor. The community includes other members who may be involved in the activity as well as the environment where the activity happens. For example, in direct observation, the community can include the learner, the subject as well as the patient, the program director or clerkship director and the location (the operating room, emergency

department or ICU). The rules that govern how the activity operates can be explicit or implicit. The division of labor refers to how the tasks necessary to carry out the activity are divided up amongst the members of the community. These three new elements, interact with and influence the subject, the object and the tools. An example of how Activity Theory can be applied to any complex activity (home/work/school balance during the COVID-19 pandemic) is found in Figure 1.

Figure 1: Application of Activity Theory to COVID Quarantine:



Why apply activity theory to the literature on direct observation?

Direct observation is designed to be an effective tool to assess learners performing clinical skills required for the safe and independent practice of medicine. Several factors beyond the rater, the learner, and the assessment tool can influence the ultimate outcome. These factors include the experience level of the rater, the learner's training level, decisions regarding who selects when and on what patient to perform the observation, the location of the observation, time allowed for the observation, the patient complexity,

and the presence of other members of the healthcare team. Because these numerous factors have the potential to impact each other, direct observation can be considered a rather complex activity system. Activity Theory enables investigators to explore the potential tensions between the extensive set of forces that are involved when a human interacts with the world around him/her. An important concept in Activity Theory is the transformation of the subject and object through their interaction which is mediated by the use of tools; which can be modified and developed through the interaction. During direct observation, there is an interaction between the rater, the learner and the object of the observation which is either a formative or summative assessment. This observation is mediated through the use of tools, such as the observation form and the patient's case complexity. Another benefit of using Activity Theory to explore direct observation is the inclusion of the community, which enables educators to consider the role that each member (e.g., the rater, learner, patient, other faculty members, program and clerkship directors, other members of the healthcare team) in the community plays. This can include power relationships between the members and the responsibilities of the community members. A unique aspect of Activity Theory is its ability to allow researchers to explore the tensions between the various elements enabling them to explore the impact that each element has on each other and ultimately the outcome. For example in Figure 1, there could be tensions between the parent and division of labor (how should he/she divide time between instruction of each child plus complete job responsibilities and household management?).

Thus, in this thesis I use Activity Theory as a guiding framework to identify and consider the numerous factors involved in direct observation and the equally important interplay between these factors. To this end, a literature review was conducted focusing on publications from the previous decade that described factors impacting direct observation of learners while engaged with real patients.

Chapter 2: **Factors and interactions influencing direct observation: A Literature Review Guided by Activity Theory** *(manuscript submitted to Teaching and Learning in Medicine)*

ABSTRACT:

Title: Factors and interactions influencing direct observation A Literature Review Guided by Activity Theory

Purpose

Ensuring that future physicians are competent to practice medicine is necessary for high quality patient care and safety. The shift towards competency based education has placed renewed emphasis on direct observation in authentic patient care contexts. Despite this interest and multiple studies focused on improving direct observation, threats to the validity and reliability of measured outcomes remain underexplored and unresolved.

Method:

We conducted a literature review of direct observation in authentic patient contexts by systematically searching databases PubMed, ERIC, Embase, and Web of Science. The authors extracted key data from the studies and applied Activity Theory to the findings to identify factors affecting these observations and the interactions between them.

Results:

19 articles were included in the analysis, 13 involved GME learners and 6 for UME. Of the 19, 6 were in the operating room setting and 4 were in the Emergency department. Using Activity Theory, we discovered that while numerous studies focus on rater and tool influences, very few study the impact of social elements. These are the rules that govern how the activity happens, the environment and members

of the community involved in the activity and how completion of the activity is divided up among the members of the community.

Conclusions:

Viewing direct observation through the lens of Activity Theory enables educators to implement curricular changes to improve direct observation of assessment. Activity Theory also allows researchers to design studies to focus on the identified underexplored interactions and influences.

BACKGROUND

Ensuring that future physicians are competent to practice medicine is necessary for high quality patient care and patient safety. For decades, direct observation of learner competence during real patient interactions has been a frequent instructional and assessment method in medical education, which has major implications for learners' progression through training (15,27) However, attention to direct observation has intensified due to recent emphasis on competency-based medical education and accreditors' recommendations to use direct observation to assess Milestones and Entrustable Professional Activities.(12,13) This new emphasis and the potential immediate and downstream effects on the learner, the profession of medicine, and patient safety, necessitates that direct observation be a reliable method of assessment. Several studies have described potential threats to the validity of direct observation, however most focus on either the rater, the learner or the tool and do not fully explore other factors and the interactions between them.

Direct observation is the process in which a supervising physician is present in the clinical environment with the learner to make a determination of learner competence. This includes assessments of the learner with a patient, during a patient hand-off, or communicating with members of a healthcare team. In most instances, the rater is conducting the direct observation based on requirements of the Accreditation Committee on Graduate Medical Education (ACGME), the program, and/or medical school. At the same

time the rater has other responsibilities including ensuring safe patient care, productivity metrics, etc. to consider during the assessment. The rater also must consider the learner's performance in the context of the patient case, which is unpredictable. In addition to providing faculty with information regarding how well the learner is progressing from novice to expert, direct observation can benefit the learner by providing opportunities to learn in authentic clinical settings, to receive more meaningful feedback, participate in learner-centered instruction, and to feel more comfortable when performing the skills being observed.(28,29)

Despite the acknowledged importance of reliable and valid direct observation, factors beyond the learner's demonstrated knowledge, skills, and attitudes, have been identified as influencing the outcome of the assessment. These factors may have implications which threaten the validity of the rater's possible conclusions. Several knowledge syntheses have investigated these factors.(19,30-33) For example, Gingerich, et al found that case complexity and rater "idiosyncrasies" were responsible for much of the variance in direct observation. (30) In another review, Tavares explored rater assessments through the lens of mental workload arguing that increasing demands on the rater "overload cognitive capacity"(31,p298) results in potential rater short-cuts and rater error. Additionally, a systematic review by Lee classified factors affecting direct observation as either intrinsic or extrinsic to the rater.(19) Intrinsic factors included, but were not limited to, training and experience level of the assessor whereas extrinsic included rater specialty, case complexity, and setting of the encounter.

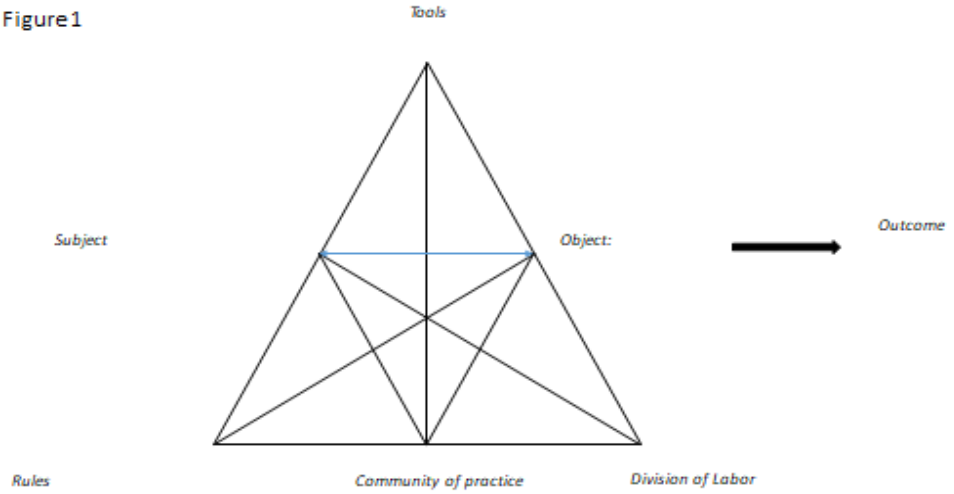
While these publications are valuable, they do not fully explore the influence of extrinsic factors, such as rules for conducting the assessment, the environment, or the amount of time and resources required. Additionally, researchers have not extensively studied how these factors interact upon and influence each other. For example, Lee excluded observation of procedural skills, and although external factors were noted, the focus was on intrinsic rater factors (19). Additionally, in another review, Gingerich focused on how raters create judgments, breaking them down into 3 separate perspectives. The third perspective, addressing rater idiosyncrasy, mentioned the importance of context in rater

assessments. Context was described as the "dynamic interaction between everyone and everything within the environment". (3,p1060) Although both reviews acknowledged factors beyond the rater and the learner, neither fully considered the interaction of all of these various factors upon each other. Due to the varied nature of these factors, along with the complexity of performing direct observation, we propose that there is value in expanding upon these reviews and in using Activity Theory (8,9) to guide our effort in unpacking the complexity of factors, their interactions, and the impact on direct observation.

Activity Theory provides a framework to understand and analyze the complex activities of humans, the systems in which they work, and the interactions or tensions between multiple factors associated with the activity.(8,9) The key concept of Activity Theory is that there is a subject (the person doing the activity) who utilizes tools (either tangible or intangible) to accomplish the activity, to work on an object (usually described as the reason or purpose of the activity) in order to achieve an outcome. In the late 1990's, Engstrom expanded on AT and included elements related to the social context of the activity.(8) These three new elements, the community; the rules that govern how the community operates; and the division of labor between the members of the community all interact with and influence the subject, the object and the tools.(Figure 1) To explore this interplay of factors, we will review the literature on direct observation using Activity Theory, which will enable us to identify relationships between factors and expose gaps. Finally, we will suggest strategies to address the tensions that arise among the factors affecting the success of direct observation.

Figure 2: Basic elements within an activity system

Figure 1



Methods:

We conducted a review of the medical literature focused on factors impacting direct observation assessments in the authentic environment of patient care. The members of the coding team included 3 clinicians (MS, LP, SW) with several years experience of performing direct observation in the clinical setting at the UME and GME level, and a non-clinician (LM), with expertise in conducting knowledge syntheses. DT, a clinician and health professions education researcher, provided guidance on the application of AT to data collection and analysis.

Data Collection

Consulting with a medical librarian and guided by our specific research question, we searched PubMed, Embase, Web of Science, and ERIC using combinations of medical subject headings and keywords including observer variance, observer bias, rater cognition, rater stringency, direct observation, workplace based assessment, and mini-CEX (34) (See Appendix A for complete search strategies). To enhance the search strategy, articles selected for full-text analysis were examined and relevant terms were added to the search. The literature search was iteratively conducted with multiple returns to the databases and

modifications of the search terms were based on the abstracts selected for review. The final search was conducted in June 2020. Reference lists of the full-text articles selected were hand searched for additional potential articles. All references were managed in EndNote.

Inclusion and Exclusion

MS and SW independently screened titles and abstracts. We included journal articles published in English between the years 2010-2020. To be included for fulltext review, studies had to be conducted in the patient care context with authentic patients, either as a live encounter or a video recording of an actual encounter, and focused on factors that impacted the direct observation of UME or GME trainees. LM served as a tiebreaker when MS and SW were not in agreement about inclusion. When a decision could not be made based on title and abstract review, it was included for full-text review. We excluded articles or published abstracts not conducted in the patient care context (e.g., OSCEs) or those involving simulation, allied health professionals, or non-UME/GME learners. Additionally, we excluded studies not involving observation of a learner by a more experienced physician for the purpose of learner formative or summative feedback. For example, observation of the validity of rating tools for making a diagnosis, or observer variability in the use of a diagnostic scale were excluded. Studies where raters watched a recorded video of a learner interacting with a standardized patient were excluded.

Data Extraction

We created a data abstraction sheet by adapting a Best Evidence in Medical Education synthesis coding sheet (35) to align with our research aim. MS and LM met on three additional occasions to refine the coding sheet for clarity. Team members independently piloted the form using a shared subset of included articles and met to discuss any discrepancies in extraction and resolve any questions regarding the extraction form. The final version of the extraction sheet was operationalized in GoogleForms.

We extracted key information (author, title, study population, type of study, type of assessment etc). We also extracted factors related to the rater (experience as a rater, specialty, gender, training on the evaluation form), the learner (training level, gender, familiarity with the rater, etc), the direct observation form (reliability and validity data, time to complete, paper versus online, etc), and where the evaluation was performed (operating room, emergency department, clinic, ward, etc) MS extracted data from all articles. LP and SW independently extracted data from 7 articles each. LM extracted data from 11 articles. MS compiled all coders' data for comparison between coders. MS and the coders met to reconcile any differences in the data extraction. Consensus was reached by discussion.

Data Analysis

After initial data extraction, to ensure mutual understanding of the six key components of AT the research team met to draft working definitions and create examples related to direct observation. During these discussions, the team agreed upon the process for how they would be applied to the included studies, including how they would be mapped onto the AT diagram (Figure 1). Next, for each study MS independently identified which components of AT were present and the relationships between the components. For a relationship to be included in the interaction analysis, the factors had to be studied and their interaction or impact on the study outcome had to be reported in the results or discussion section of the article. MS then met individually with LP, LM and SW to confirm this AT-focused coding for those studies that they had independently coded earlier and to resolve any discrepancies via discussion.

Results:

The initial search resulted in 1,972 studies of which 19 were included.(5,6,21,36-51) (Figure 2). Summary characteristics of each article are listed in Table 1.

Figure 3: Literature review flow chart

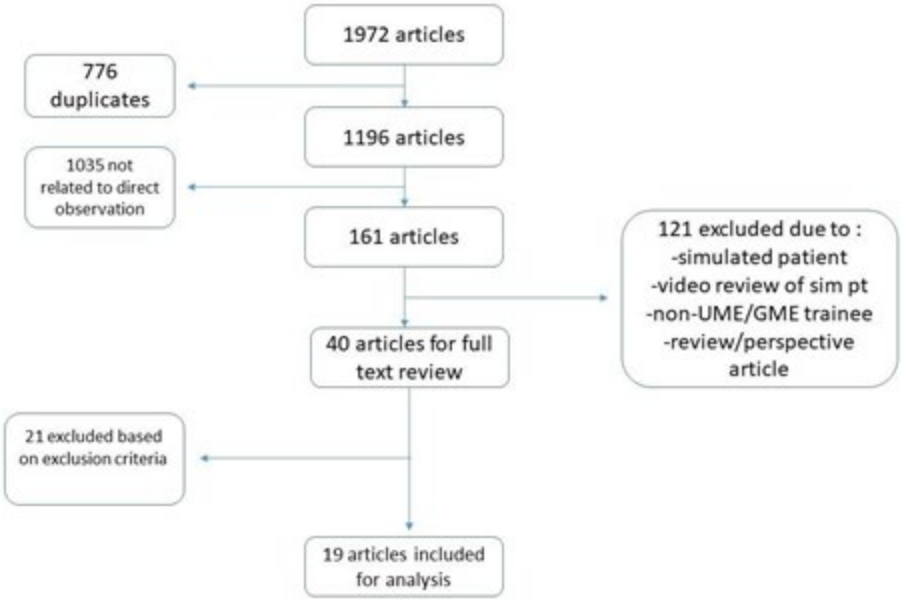


Table 1: Summary characteristics of included articles (n=19)

	Characteristics	N (%)
Country of origin	United States	8 (42.1)
	Canada	3(15.7)
	Netherlands	3(15.7)
	Australia/New Zealand	3(15.7)
	Taiwan	1(5.3)
	United Kingdom	
Learner level	Graduate Medical Education	13 (68.4%)
	Undergraduate Medical Education	6 (31.6)

Location of observation	Operating room	6 (31.6)
	Emergency room	4 (21)
	Other	9 (47.4)
Rater	Faculty	19 (100)
	Non-physician	2 (10.5)
	Resident	1 (5)
Observation form	Mini-CEX	7 (36.8)
	Workplace based assessment	3 (15.7)
	Zwisch Autonomy Scale	2 (10.5)
	OPRS	1 (5)
	Direct encounter card	1 (5)
	GOALS	1 (5)
	T-MEX	1 (5)
	Not specified	3 (15.7)
Intent of observation	Formative	8 (42.1)
	Summative	2 (10.5)
	Not specified	9 (47.4)
Method of observation	Live patient encounters (n=15)	15 (78.9)
	Video review of real patients (n=4)	4 (21.1)

Activity Theory analysis

All of the included studies described factors, which we were able to map to the elements of AT. In the following paragraphs, we list the factors in relation to the AT elements and then describe the observed interactions between these elements.

Subject (n=19)

The subject included the rater and the associated traits that he/she brought with him/her to the observation. Factors mapped onto the subject element were the rater's level of experience performing direct observation (n=9),(6,21,36,37,41,42,45,48,50) training on the tool (n=6),(36,40,41,44,48,51) specialty, (n=2) (6,37) and rater gender (n=3).(21,37,38) For example in Chang, raters were classified by gender, specialty (Emergency Medicine versus Trauma surgeon), and level of experience in performing direct observation to determine if these characteristics had an impact on direct observation.(37)

Tools (n=19)

The subject uses tools to perform the action. Tools were classified as things the subject considers when performing the direct observation of the learner, whether tangible (n=19) or intangible (n=9). Tangible elements included the assessment form (n=16),(6,21,36-40,43-51) familiarity with the learner (n=5),(6,38,42,44,45) and learner level of training (n=5).(21,38,40,42,51) Intangible components were learner behavior (n=5),(21,39,41,42,44) patient behavior (n=2),(41,42) and patient/surgical case complexity (n=9).(5,6,36,41-46) Ten articles specifically examined the impact of the direct observation form on study outcome.(6,36-39,40,45-47,49,50). Familiarity with the learner was mentioned in 5 studies.(6,38,42,44,45) The complexity of the patient's disorder or surgical procedure was noted in 7 studies.(5,21,36,42-44,46) In one study, raters were grouped by familiarity with the residents, noting that there was a potential rater bias based on knowing the level of training of the surgical resident and familiarity with their prior surgical performance.(6)

Object (n=19)

The object, which is the purpose of the activity system, was defined in our analysis as the act of assessing or observing the learner. There were 6 studies dedicated to procedural assessment (5 surgical residents and 1 anesthesia residents). Direct observation in eight studies assessed overall clinical performance, of which three also included whether or not the learner received feedback and/or an action plan.(37,40,47) Olupeliyawa examined observation of medical students working as part of a healthcare team and assessed their communication skills, recognizing their individual limitations, and including other healthcare members in the plan of care.(47)

Rules (n=12)

Rules included things that govern how the subject interacts with the community and the tools in order to influence the object, as well as things that impact how the division of labor happens. In our analysis this included whether the study specified rules about the provision of feedback (n= 3),(37,40,47) who selected when the observation happened (n=5)(39,40,46,47,49) and requirements regarding the timing of form completion (n= 4).(21,40,49,51) Donato provided monetary incentives for completing direct observation weekly, and faculty were allowed to select the setting; however the form must have included an action plan (40) Two studies examined the quality of completion between computer and paper-based forms, both finding the completion rate was higher in the online format,(37,49) however only 1 found the quality of the assessment improved based on the format used.(37) Multiple studies examined the number of assessments needed for an accurate judgement regarding trainee performance (n=5).(39,43,47,48,51)

Community (n=19)

Community is defined as the members of the social circles that are affected by and also influence the subject, the rules, and the tools as the action is being carried out. The community also influences the division of labor. We included learners (n=19) and faculty members of the training programs or schools (n=1). We also included the environment where the observation occurred (n=10),(5,6,21,36-38,44,49-51) (eg,the medical school (n=6),(5,41,42,44-46) and the residency training program (n=14)) as part of

the community. An article by Cheung compared direct observation of three groups within the hospital community: Emergency Medicine residents who had longitudinal relationships with the supervisor, EM residents who were not part of a continuity team, and residents who were from different residency programs, but required to rotate in the Emergency Department.(38)

Division of Labor (n=8)

Division of labor is the process in which the actions occurring in the activity system are distributed between the subject and community, while being mediated by the rules. In this analysis, the decisions regarding when to be the teacher, the assessor, or the physician responsible for patient care were included in this category. We also included the time needed to complete the observation and time for provision of feedback (n=3)(40,46,47) as well as the number of assessments needed for a reliable outcome (n=6).(36,39,43,47,50,51) For example, the study by Donato included the time required to complete the form, the cost of printing the form, and the time required to record the comments.(40) Chang discovered that the amount of feedback comments included on the form was related to how busy the Emergency Department was on that particular shift. (37)

Interactions among elements:

Beyond classifying factors in relation to the elements of AT, we used AT to examine the articles for evidence of relationships and interactions between the 6 elements. We discovered varying degrees of interaction which could impact the outcome of the activity system.

The most common interaction, identified in all 19 articles, involved relationships among the subject, tool, and object. Within this interaction, the tool included the direct observation form, the patient or surgical case complexity, the learner's level of training, and familiarity with the learner. In this situation, the subject and tool influenced each other and the object. Lane discovered "Rater gender was independently associated with meaningful autonomy, as both female and male residents were less likely to receive

meaningful autonomy from female raters." (21,p242) The subject element included the rater and their gender, the tool was the learner's gender, and the object was an assessment of the learner's meaningful autonomy in the operating room. The interaction influenced the ultimate outcome with female raters being less likely to give meaningful autonomy to residents regardless of gender. Another study examined whether familiarity bias could account for poor inter-rater reliability between local faculty and American Board of Neurology and Psychiatry (ABPN) examiners. In this study the subjects are the ABPN external faculty, faculty from the residents training program, and faculty from another local university. The experience of the rater in assessing was also a subject element. The tools were the Neurology Evaluation Exercise and familiarity with the resident. The object was determination of an acceptable or unacceptable score on the assessment. The authors concluded that the poor correlation between these groups was not the result of familiarity bias, suggesting that the tool (the Neurology Evaluation Exercise), and ABPN examiner experience were potential factors.(45) Both the subject elements and the tool elements impacted the object element which then influenced the outcome; poor correlation between raters viewing the same performance.

Interactions between the subject, tool, division of labor, and the object (n=7) were also prevalent. Most of these relationships centered around the number of observations required to improve the reliability of the assessment. For example, Olupeliyawa found that "eight T-MEX forms is sufficient for all the different clinical situations that the T-MEX targets. Students found completing 3-6 T-MEX encounters to be feasible in a six-week term." (47,p364) Although these relationships were mentioned, the studies did not address the impact of this number on the rater, the learner, or the community. One study did specifically mention role tension between faculty roles as teacher and assessor. "Major factors were 'clarity', overall impression of the trainee, role tension especially when giving constructive feedback, prior knowledge of the trainee, and case complexity."(44 p89)

Although rules regarding the direct observation were mentioned in 63% of the studies, only four (21%) detailed an interaction or relationship between the rules and other elements of the AT

diagram.(37,40,44,47) Chang, et al specified that all three aspects of feedback were to be completed on each mini-CEX (rule), however they discovered that more experienced raters provided all three aspects of feedback only 8.6% of the time, compared to 37.1% for junior emergency physicians.(37) Clarity of what was being evaluated and "institutional expectations relating to the assessment"(44 p92) were specifically mentioned as points of contention by faculty members in Lee's study.

Few studies (15.7%) acknowledged the impact of the environment, which we classified as part of the community, on direct observation. Chang considered how busy the emergency department was as a potential reason for the lack of complete feedback on all mini-CEX forms, concluding that raters spent less time providing feedback to learners and less time completing the form when the emergency department was more busy. Therefore busyness of the emergency department was influencing the subject, the division of labor, the community and the object.(37) Kelleher considered raters observing residents in different contexts as one of the sources of rater variance,(43) citing entrustability varies based on whether the observation is happening in the intensive care unit, a ward or clinic. Thus highlighting the influence of the community on the subject and the object. Lee identified busyness of the emergency department, interruptions, and background noise as factors that influence rater judgement, noting that busy assessors may rely on gut feelings and more easily measured competencies when pressured for time.(44)

Discussion

In this study, we identified and described the key factors influencing direct observation. Going a step further, we applied Activity Theory as a guiding framework for considering the interplay of these factors. This application illuminated the interacting elements of the direct observation activity system such that underexplored tensions and influences were discovered and will be discussed below.

Several studies have analyzed the influence of rater cognition and rater experience on direct observation.(33,41,42,52-55) Numerous studies, including several Generalizability Theory analyses, have focused on the reliability and validity of the outcomes measured when using a wide variety of tools including the

mini-CEX.(5,39,43,48,50,51) The application of Activity Theory to direct observation enabled us to identify a gap in the literature regarding the collective influence of rules, community, and division of labor on this complex activity. These factors are key elements of the social aspect of the activity, which are frequently overlooked in studies regarding direct observation.

The impact and influences of community elements including the patient and the environment were not studied in the majority of the reviewed literature. Beyond identifying the role of each individual in the room and informing the patient that in addition to receiving care, they are participating in assessment of a learner; Activity Theory allows educators to consider other potential tensions involving the patient. Several of the studies in this analysis noted a role of patient complexity, however the full effect of this on the outcome of the assessment was not considered. Patient complexity and patient behavior (a more talkative or cognitively impaired patient) may impact division of labor with a more complex patient potentially requiring additional time for observation necessitating adjustments to faculty and learner schedules. Feedback to the learner could then be impacted due to additional time required for observation. Through the lens of Activity Theory, these potential tensions involving the rules, community and division of labor can be considered and addressed by educators.

The particular environment where the observation happens is another significant component of the activity system. An assessment of a learner in a busy emergency department presents challenges that are different than what is encountered in an assessment of a patient presenting with a routine problem in a clinical setting. Several of the studies included in our analysis identified an impact of patient or surgical case complexity; however only 3 acknowledged the effect of the environment on the assessment. Some of these issues were mentioned in a Guideline published by Kogan, and we feel that Activity Theory provides a framework for considering how to implement the “do’s”, how to mitigate the “don’ts”, and possibly enable research to answer the “don’t knows” of direct observation.(26) For example, two of the “don’ts” suggested for educators can be addressed through Activity Theory. Kogan recommended that the responsibility of asking for observation should be shared between the learner and the rater.

Additionally, she recommended attention to the "tension between being both a teacher and assessor" (p297). Activity Theory enables educators to develop rules addressing when and how to conduct direct observation factoring in the roles of community (environment, patient, other faculty) and the division of labor between members of the community (ensuring adequate uninterrupted time for observation and provision of feedback). (Figure 3)

Beyond being a useful lens to view the assessment process of direct observation, Activity Theory can assist in curricular development centered on competency based medical education. During curricular design, Activity Theory enables programs to not only think about stakeholders, resources, time and faculty requirements, background of the learner, needs assessment, and the method of evaluation, but to also assess the influences and tensions between and among these various elements. The numerous components of this complex activity system do not act in isolation, nor do they proceed in series. When designing competency based curricula incorporating direct observation it is important to clarify the roles of the community including the patient and environment and develop rules that address the division of labor between the community and the subject. Additionally, examining rater and learner cognition, identifying the impact of patient complexity and exploring the influence of the tool are essential elements when designing a curriculum centered on direct observation.

Limitations

This study must be considered in light of its limitations. It is possible that despite our best efforts, we may have inadvertently missed a study. However, by searching multiple databases and hand searching the reference lists of the related reviews, we feel that we attempted to address this potential issue. We chose to focus on only literature related to medical education, therefore it is possible that the gaps and tensions we identified have been reported in other studies outside this context (e.g., factors that impact direct observation of nursing students.) By selecting AT, we took a stance, which we felt is warranted, but this lens, like all theories, means not illuminating certain aspects that another theory may have.(58)

Conclusions:

In the complex activity system of direct observation of trainees' readiness for independent practice we employed AT in a review of prior publications to identify systematically factors, which influence the observation, and to locate these within the dimensions of AT - the rater, the tool, object, the community, the rules and division of labor. We paid particular attention to the relationships and tensions between these six factors of AT, and identified areas for consideration and exploration by educators, including curricular development and gaps for further study.

Chapter 3: Discussion and military relevance

This thesis synthesizes the literature to describe what is known about the factors influencing direct observation in direct patient care contexts and uses Activity Theory to illuminate how all of these factors influence each other. This literature review presents two very important findings. First, most of the literature emphasized unidirectional influences (i.e., the influence of the rater on the outcome or the influence of the tool on the outcome). Second, the social aspects of the activity system were relatively unexplored in the majority of studies reviewed. How direct observation happens, who it affects (learners, faculty, program directors, patients, etc.), and the time required to observe, assess, and instruct are essential elements to consider when designing a curriculum utilizing direct observation. Using Activity Theory enables consideration of the multi-directional impact of each element within the direct observation activity system and the influence of the social elements, potentially enabling educators and researchers to approach direct observation in a new light.

One benefit of using Activity Theory is identifying the impact of the components on each other i.e., a "multi-directional" interaction. However, the majority of interactions identified in our analysis were only described in the literature as "one-way"; usually the impact of rater factors on the observation. This is a lost opportunity because it neglects the impact of non-rater factors on each other and the impact of those interactions on the rater. For example, how does the patient complexity impact the learner's performance or is there an effect of the learner's training level on patient behavior? Other examples include the influence of a summative evaluation on the rules of the direct observation (is the patient, the timing, and the environment chosen by the learner or the rater?). The impact of rater experience, rater training, and rater cognition are necessary to consider when conducting direct observation, however using Activity Theory, education researchers can also explore the influence of multiple elements involved in the activity system.

The majority of the articles included in this knowledge synthesis studied the effect of the subject and tool on the observation and the impact these two factors may have on the validity of the measured

outcome. For example, London, et al studied the impact of internal versus external faculty on neurology resident pass rates during a direct observation summative exercise. Results indicated poor inter-rater reliability with more experienced external faculty being more likely to fail a resident than local faculty. (45) However, the impact of the observation on the rater and learner should also be considered when studying competency based assessments. For example, does the rater knowing they are going to conduct a direct observation with a learner affect how they conduct clinical care and does it change their behavior towards this learner? The interactive nature of Activity Theory, enables educators to also evaluate what is the effect of being observed on the learner. LaDonna et al conducted an interview study with 22 learners about their experiences and perceptions of being observed and concluded that learners frequently altered their behavior to please the observer (observer centered care) rather than providing patient centered care.(7) Thinking about direct observation through the lens of Activity Theory, raters and learners could more thoughtfully consider the implications of the observation on the encounter and the relationships among the rater, learner and patient; both community factors.

Examining the process of direct observation through Activity Theory could enable education leaders at both UME and GME programs to more systematically consider multiple elements of this complex activity. Residency program and medical school curricular design may also benefit when applying Activity Theory to direct observation; in particular when considering the lower portion of the Activity Theory triangle (see figure 1) that contains the social elements of the activity system -the rules, division of labor, and the community. For example, if a generalizability study discovers that eight assessments are necessary to increase reliability of the outcome, the residency or clerkship must adapt to ensure that each learner is able to be assessed eight times. While this information is valuable these eight assessments would not happen in isolation, which has implications for planning. Education leaders must also consider over what time period, and when will the observations happen. The residency and clerkship would need to create rules governing how often the assessments are done, the program or

clerkship director (community members) will need to determine if more faculty are needed and will need to ensure sufficient time and training for faculty to conduct the observation (division of labor).

This review also highlighted gaps in the literature. For example, few studies examined the impact of gender and ethnicity on direct observation. Research examining the influence of gender on the OSCE have found that the gender of the rater, the learner and of the standardized patient may impact the outcome.(22, 56-60) Only a single study specifically examined the influence of gender on direct observation. This study found that female raters were less likely to provide meaningful autonomy to trainees regardless of gender. (21). This review did not find any studies that included or examined rater or learner race/ethnicity during the direct observation assessment despite literature supporting that medical students from under-represented minorities receive lower clerkship scores and receive different narrative comments on performance evaluations. (61-63) These findings may affect competitiveness for residency selection and ultimately career path for learners. Rater gender and ethnicity are attributes that should be considered under subject factors, learner ethnicity and gender may be tools used by the rater in completing the observation and comparisons between an under-represented minority student with white counterparts could be considered under community influences. An approach to direct observation using Activity Theory could enable educators to consider these influences and interactions at a more conscious level, with the goal of mitigating differences based on factors other than learner performance.

Clarifying the roles of each member of the community, developing rules necessary to conduct the observation, considering rater and learner cognition, identifying the impact of patient complexity and environment, examining the influence of the tool, and establishing protected time for the observation are all essential when thinking about competency based assessment using direct observation. Recognizing that potential barriers to complete objectivity in direct observation will always exist, the application of Activity Theory to this activity enables all members of the community to more carefully consider the factors that can influence direct observation and design research studies to examine them in more detail.

Military Relevance:

Graduating competent physicians to take care of the armed forces and their families is part of the Uniformed Services University (USU) and military Graduate Medical Education mission statements. As educators, we need to ensure the quality of our graduates not only meets but hopefully exceeds the standards established by the residency programs and the school. Integral to this is ensuring that decisions regarding whether or not a learner has met the established competencies for advancement and independent practice are predominantly, if not exclusively, related to the knowledge, skills, and attitudes of the learner. Educators need to reduce the impact of external influences.

The National Defense Authorization Act proposes staffing reductions at many of the military training facilities as well as elimination of some training programs (64). This will undoubtedly have an effect on the ability of programs and faculty to provide high quality direct observation to learners. An analysis of the direct observation process through Activity Theory enables education leaders at the Military Treatment Facilities (MTF) and at USU to identify requirements for carrying out this complex activity. Program directors, clerkship directors and faculty members can develop their curricula factoring in rules, community and division of labor requirements using Activity Theory as a guide. Consideration of the community elements is potentially even more important in military medical education. Unlike our civilian counterparts who have more power to select exactly who becomes a member of the university or program teaching faculty, physicians assigned to a MTF are generally selected by leaders who may not understand the educational needs of clerkship and residency programs. Faculty at the MTF may have limited to no experience in direct observation or medical education and may have a different understanding of their role within the department. This represents somewhat unique considerations for clerkship and residency directors and are reflective of the subject, rules, and community elements that Activity Theory explores. Understanding the complexity and value of direct observation through this lens may enable program directors to have more in-depth discussions with department heads and hospital leadership regarding dedicated time for teaching and the utilization of other productivity metrics beyond pure patient volume seen per day.

Conclusion

Identification of potential factors that influence rater assessment is essential to improving the objectivity of direct observation, enhancing the quality of patient care, and improving the observation skills of future military physicians who will someday become assessors themselves. This thesis lays the groundwork for understanding these factors and their interplay. Additionally, it provides a new approach for investigating some of the challenges of performing direct observation and the perceived lack of objectivity of these assessments. This thesis exposed some important gaps in direct observation research, mainly the impact of the social elements on the process and the bidirectionality of each element in the activity. The influence of rater and learner gender on direct observation is another under-explored area. The impact of patient gender on the outcome is also unknown and could be another area for future research.

References:

1. Aubin AS, St-Onge C, Renaud JS. Detecting rater bias using a person-fit statistic: a Monte Carlo simulation study. Perspect Med Educ 2018;7:83-92

2. Holmboe ES. Faculty and the observation of trainees' clinical skills: problems and opportunities. Acad Med 2004;79:16-22

3. Gingerich A, Kogan, J, Yeates P, et al. Seeing the 'black box' differently: assessor cognition from three research perspectives. Medical Education 2014:48:1055-1068.

4. Cheung WJ, Patey AM, Frank JR, et al. Barriers and enablers to direct observation of trainees' clinical performance: a qualitative study using the theoretical domains framework. Acad Med 2019;94:101-114

5. Krieter CD, Wilson AB, Humbert AJ, Wade PA. Examining rater and occasion influences in observational assessments obtained from within the clinical environment. Med Educ Online 2016,21:29279 http://dx.doi.org/10.3402/meo.v21.29279

6. Bjorklund KA, Sommer N, Neumeister M, et al. Establishing validity evidence for an operative performance rating system for plastic surgery residents. J Surgical Education 2018;76:529-539

7. La Donna K, Hatala R, Lingard L, et al. Staging a performance: learners' perceptions about direct observation during residency. Medical Education 2017 doi: 10.1111/medu.13232

8. Engstrom Y. (1999) Activity theory and individual and social transformation. In Engstrom Y, Miettinen R, Punamaki R. Eds. Perspectives on Activity Theory. pp 19-38. Cambridge University Press

9. Hashim, N. H. and Jones, M. L.: Activity Theory: A framework for qualitative analysis 2007 http://ro.uow.edu.au/commpapers/408

10. Ten Cate O, Billett S. Competency-based medical education: origins, perspectives, and potentialities. Med Educ 2014;48:325-332.

11. Pangaro L, ten Cate O. Frameworks for learner assessment in medicine: AMEE Guide No. 78. Med Teach 2013;35:1197-1210

12. Edgar L, Roberts S, Holmboe E. Milestones 2.0: A step forward. J GME June 2018 DOI: http://dx.doi.org/10.4300/JGME-D-18-00372.1

13. Association of American Medical Colleges (AAMC). . Core Entrustable Professional Activities for Entering Residency. 2013 https://www.mededportal.org/icollaborative/resource/887. Accessed Feb 3, 2020

14. Miller GE. The assessment of clinical skills/competence/performance Acad Med 1990;65(9 suppl):S63-S67

15. Fromme HB, Karani R, Downing SM. Direct observation in medical education: review of the literature and evidence for validity. Mt Sinai J Med 2009;76:365-371

16. ten Cate O. Entrustability of professional activities and competency-based training. Med Educ 2005;39:1176-1177.

17. Pangaro L. A new vocabulary and other innovations for improving descriptive in-training evaluations. Acad Med 1999;74:1203-1207

18. Kogan JR, Conforti L, Bernabeo E, et al. Opening the black box of clinical skills assessment via observation: a conceptual model. Med Educ 2011;45:1048-1060.

19. Lee V, Brain K, Martin J. Factors influencing mini-CEX rater judgements and their practical implications: a systematic literature review. Acad Med 2017;92:880-887

20. Downing SM. Threats to the validity of clinical teaching assessments: what about rater error? Med Educ 2005;39(4):353–5

21. Lane SM, Young KA, Hayek SA, et al. Meaningful autonomy in general surgery training; Exploring for gender bias. American J Surg 2020; 219:240-244

22. Schleicher I, Leitner K, Juenger J, et al. Examiner effect on the objective structured clinical exam- a study at five medical schools. BMC Medical Education 2017 17:71 DOI 10.1186/s12909-017-0908-1

23. Dayal A, O'Connor DM, Qadri U, Arora V. Comparison of male vs female resident milestone evaluations by faculty during emergency medical residency training. JAMA Internal Medicine 2017;177:651-657

24. Fazio, SB, Torre DM, DeFer TM. Grading Practices and Distributions Across Internal Medicine Clerkships, Teaching and Learning in Medicine 2016; 28:286-292

25. Holmboe ES, Hawkins RE, Huot SJ. Effect of training in direct observation of medical residents' clinical competence: a randomized trial. Ann Intern Med. 2004;140:874–81.

26. Kogan JR, Hatala R, Hauer KE, Holmboe ES. Guidelines: The do's, don'ts and don't knows of direct observation of clinical skills in medical education. Perspect Med Educ 2017;6:286-305

27. Anonymous. The spirit of apprenticeship in medicine. *JAMA.* 1924;83(17):1337-1338. doi:10.1001/jama.1924.02660170053021

28. Kuo AK, Irby DI, Loeser H. Does direct observation improve medical students' clerkship experiences? Med Educ. 2005;39(5):518.

29. Howley LD, 'William WG. Direct observation of students during clerkship rotations: a multiyear descriptive study. Acad Med 2004; 79: 276-280.

30. Gingerich A, Regehr G, Eva K. Rater-based Assessments as Social Judgments: Rethinking the Etiology of Rater Errors. Acad Med 2011 doi:10.1097/ACM.0b013e31822a6cf8.

31. Tavares W, Eva K. Exploring the impact of mental workload on rater-based assessments. Adv in Health Sci Educ 2013;18:291-303

32. Kogan JR, Holmboe ES, Hauer KE. Tools for direct observation and assessment of clinical skills of medical trainees: a systematic review. Jama. 2009 Sep 23;302(12):1316-26.

33. Kogan JR, Conforti L, Bernabeo E, Iobst W, Holmboe E. Opening the black box of clinical skills assessment via observation: a conceptual model. Medical education. 2011 Oct;45(10):1048-60.

34. Norcini J, Blank L, Duffy D, Fortna G. The mini-CEX: a method for assessing clinical skills. Ann Intern Med 2003;138:476-81

35. Cook DA, West CP. Conducting Systematic Reviews in Medical Education: A Stepwise Approach. Med Educ 2012 Oct;46(10):943-52

36. Bilgic E, Watanabe Y, McKendy K, et al. Reliable assessment of operative performance. Am J Surgery 2016;211:426-430

37. Chang Y, Lee C, Chen C, et al. Exploring the influence of gender, seniority and specialty on paper and computer-based feedback provision during mini-CEX assessments in a busy emergency department. Adv in Health Sci Educ 2017;22:57-57

38. Cheung WJ, Dudek NL, Wood TJ, Frank JR. Supervisor-trainee continuity and the quality of work-based assessments. Medical Education 2017;51:1260-1268

39. Crossley J, Johnson G, Booth J, Wade W. Good questions, good answers: construct alignment improves the performance of workplace-based assessment scales. Medical Education 2011;45:560-569

40. Donato AA, Park YS, George DL, et al. Validity and feasibility of the minicard direct observation tool in 1 training program. J Graduate Medical Education DOI:http://dx.doi.org/10.4300/JGME-D-14-00532.1

41. Govaerts MJB, Shuwirth LWT, Van der Vleuten CPM, Muijtjens AMM. Workplace-based assessment: effects of rater expertise. Adv in Health Sci Educ 2011;16:151-165

42. Govaerts MJB, Van de Wiel MWJ, Cchuwirth LWT, et al. Workplace-based assessment: raters' performance theories and constructs. Adv in Health Sci Educ 2013;18:375-396

43. Kelleher M, Kinnear B, Sall D, et al. A reliability analysis of entrustment-derived workplace-based assessments Acad Med 2020;95:616-622

44. Lee V, Brain K, Martin J. From opening the 'black box' to looking behind the curtain: cognition and context in assessor-based judgements. Adv in Health Sci Educ 2019;24:85-102

45. London Z, Schuh L, Gelb D, Schultz L. Education research: unsatisfactory NEX rating correlations, searching for the reasons. Neurology 2013;80:e142-e145

46. Meresh E, Daniels D, Sharma A, et al. Review of mini-clinical evaluation exercise (mini-CEX) in a psychiatry clerkship. Advances in Medical Education and Practice 2018;9:279-283

47. Olupeliyawa AM, O'Sullivan AJ, Hughes C, Balasooriya CD. The Teamwork Mini-Clinical evaluation exercise (T-MEX): a workplace-based assessment focusing on collaborative competencies in health care. Acad Med 2014;89:359-365

48. Reinders ME, Blankenstein AH, van Marwijk HW, et al. Reliability of consultation skills assessments using standardized versus real patients. Medical Educatio 2011;45:578-584

49. Smith N, Harnett J, Furey A. Evaluating the reliability of surgical assessment methods in an orthopedic residency program. Can J Surg 2015 DOI: 10:1503/cjs.010614

50. Weller JM, Misur M, Nicolson S, et al. Can I leave the theatre? A key to more reliable workplace-based assessment. British Journal of Anaesthesia 2014;112:1083-1091

51. Williams RG, Swanson DB, Fryer JP, et al. How many observations are needed to assess a surgical trainee's state of operative competency? Ann SUrgery 2019;269:377-382

52. Yeates P, O'Neill P, Mann K, Eva K. Seeing the same thing differently. Mechanisms that contribute to assessor differences in directly-observed performance assessments. Adv in Health Sci Educ 2013;18:325-341

53. Gomez-Garibello C, Young M. Emotions and assessment: considerations for rater-based judgements of entrustment. Medical Education 2018:52:254-262

54. St-Onge C, Levesque A, Varpio L. Expectations, observations, and the cognitive processes that bind them: expert assessment of examinee performance.

55. Wood,TJ. Exploring the role of first impressions in rater-based assessments. Adv Helath Sci Educ Theory Pract 2014;19:409-427.

56. Berg K, Blatt B, Lopreiato J, et al. Standardized patient assessment of medical student empathy: ethnicity and gender effects in a multi-institutional study. Acad Med 2015;90:105-111.

57. Carson J, Peets A, Grant V, McLaughlin K. The effect of gender interactions on students' physical examination ratings in Objective Structured Clinical Examination Stations. Acad Med 2010;85:1772-1776

58. Humphrey-Murto S, Touchie C, Wood T, Smee S. Does the gender of the standardized patient influence candidate performance in an objective structured clinical examination? Medical Education 2009;43:521-525.

59. Chong L, Taylor S, Haywood M, et al. The sights and insights of examiners in objective structured clinical examinations. J Educ Eval Health Prof 2017; 14: 34 • https://doi.org/10.3352/jeehp.2017.14.34

60. Mortsiefer A, Karger A, Rotthoff T, et al. Examiner characteristics and interrater reliability in a communication OSCE. Patient Education and Counseling 2017;100:1230-1234

61. Lee K, Vaishnavi S, Lau S, et al. "Making the grade:" noncognitive predictors of medical students' clinical clerkship grades. J National Medical Association 2007;99:1138-1150

62. Rojek A, Khanna R, Yim J, et al. Differences in narrative language in evaluations of medical students by gender and under-represented minority status. J Gen Intern Med 2019;34:684-691

63. Low D, Pollack S, Liao Z, et al. Racial/Ethnic disparities in clinical grading in medical school. Teaching and Learning in Medicine 31:5, 487-496, DOI:10.1080/10401334.2019.1597724

64. H.R.2500 - National Defense Authorization Act for Fiscal Year 2020. https://www.congress.gov/bill/116th-congress/house-bill/2500 (accessed 12/29/2020)