Semantic Competence as a Marker of Clinical Reasoning Performance

By

Michael Berge, MD

Thesis submitted to the Faculty of the Health Professions Education Graduate Program, Center for Health Professions Education, 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.

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

August 2, 2021

APPROVAL SHEET

Title of Thesis: Semantic Competence as a Marker of Clinical Reasoning Performance

Name of Candidate: Michael Berge, MD

08/02/2021

THESIS AND ABSTRACT APPROVED:

DATE:

[redacted]

8/2/21

_______________________________ _________
Ronald M. Cervero, PhD
CENTER FOR HEALTH PROFESSIONS EDUCATION
Committee Chairperson

[redacted]

___8/2/21_____

Steven J. Durning, MD, PhD
CENTER FOR HEALTH PROFESSIONS EDUCATION
Thesis Advisor

[redacted]

8/2/21

Michael Soh, PhD
CENTER FOR HEALTH PROFESSIONS EDUCATION
Thesis Advisor

ACKNOWLEDGEMENTS

Throughout my Health Professions Education journey, I have enjoyed the excellent mentorship of numerous instructors, colleagues, and advisors. As with everyone around the world, the last year has seen immense trauma, challenge, change, and growth. Through it all, I have grown immensely as a leader and as an educator. In particular, my research advisors Drs. Michael Soh and Steven Durning have been immeasurably patient with my progress, often stalling in the face of personal and professional demands. Only through their strong support, leadership, and mentorship have I been able to succeed in this endeavor. I only hope that my continued work in the field of Health Professions Education can bring honor to the exceptional lengths to which they have gone to give me this opportunity.

DEDICATION

This thesis is dedicated to my wife, Megan, and our children Eleanor, Caleb, and Amelia. They have been very supportive as I have pursued this personal and professional growth opportunity and without their love and sacrifice, this would not have been possible.

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.

COPYRIGHT STATEMENT

ABSTRACT

Semantic Competency as a Marker of Clinical Reasoning Performance

Michael Berge, MD, 2021

Thesis directed by Steven Durning, MD, PhD, and Michael Soh, PhD, USUHS Center for Health Professions Education

Purpose: This thesis sought to explore the relationship between semantic competence (or dyscompetence) displayed during think alouds performed by resident and attending physicians and clinical reasoning performance.

Methods: Internal medicine resident physicians and practicing internists participated in "think-alouds" after watching videos of typical presentations of common diseases in internal medicine. The think-alouds were transcribed and evaluated for the presence of semantic competence (and dyscompetence) and these results were correlated with clinical reasoning performance.

Results: We found that the length of think-aloud was negatively correlated with clinical reasoning performance. Beyond this finding, however, we did not find any other significant correlations between semantic competence (or dyscompetence) and clinical reasoning performance.

Conclusion: While this study did not produce several of our hypotheses of correlation between semantic competence and clinical reasoning performance, we discuss the possible implications and areas of future study regarding the relationship between semantic competency and clinical reasoning performance.

LIST OF TABLES

Table of Contents

CHAPTER 1: Introduction

Background

The goal of most professional education programs, both within healthcare and other fields, is to achieve *expert performance.* Expert performance is achieved through deliberate practice, which requires effortful practice to not only the performance of a task but also reflective dedication to *improving* that performance (Ericsson, 1993). To do this well, both the educator and the learner should ideally have an arsenal of tools at their disposal to aid them in this critical reflective task, including methods to assess the underlying processes that lead to successful performance.

Clinical reasoning can be defined as the task by which the health professional gathers and processes clinical and contextual information and uses medical knowledge to make patient care decisions, encompassing all the steps up to and including the arrival at a diagnosis or management plan for a patient (Croskerry, 2009). This task is certainly one of the central pillars of physician performance in any specialty and therefore warrants considerable focus in our health professions education programs; perhaps more than any other subject, health professions training should strive to produce clinical reasoning expert performance.

Deliberate practice toward achieving expert clinical reasoning performance requires that we find ways to assess not just the outcome of clinical reasoning (e.g. diagnosis and management plans), but also underlying processes that are occurring within the trainee's mind (e.g. steps proximal to the diagnosis). While there are methods to assess clinical reasoning, it remains challenging, particularly evaluating the processes underlying clinical reasoning (Daniel, 2019). Illuminating these methods necessitates a robust understanding of clinician thought processes, particularly the concept of knowledge organization, which refers to how information is stored in the mind so that it can be rapidly and effectively recalled and used in practice. Semantic competence, which is the use of specific abstract terms to describe a patient's presentation, is believed to be an early transformation of knowledge organization.

Prior research by Bordage et al has pointed to the use of semantic qualifiers, or specific demonstrations of semantic competence, during case presentations as a potential marker of more successful clinical reasoning (meaning that the subject is more likely to make the correct clinical decisions and consider the correct diagnoses in their differential) (Bordage, 2007), however this is an area in need of further study. The aim of this thesis is to further evaluate the link between semantic competence and clinical reasoning performance, with the belief that the former may be a useful marker of underlying knowledge organization.

Theoretical Frameworks

This thesis is informed by several theoretical frameworks, chief among them are illness script theory, cognitive load theory, and deliberate practice. Also, the concepts of knowledge organization which plays a key role in the clinical reasoning process, scaffolded this study. These factors are all at play during the clinical reasoning process and displayed to varying degrees as part of the think aloud process. While some of these theoretical frameworks were not explicitly manifest in our research study, having a thorough understanding of these frameworks is important for placing our research within the larger context of understanding the processes underlying clinical reasoning performance.

Illness script theory holds that information about a specific disease or pathology is stored in long term memory with varying degree of accessibility and inter-connectedness based on that clinician's experiences (i.e. a clinician who has seen many patients with a specific disease in many different contexts and presentations may have a better-developed illness script for that disease) (Custers, 2015). A novice clinician may have a more difficult time accessing the illness scripts in their long-term memory due to less development and/or interconnectedness, if such scripts exist at all, particularly for those who have either not seen a particular illness or who have not yet established well-organized knowledge on the topic in their long-term memory (ten Cate, 2018).

It follows, therefore, that understanding human memory is also essential to this thesis. Long-standing theory of human memory describes the process by which sensory information is taken in and briefly retained within working memory. Working memory is limited in both time and capacity. Long-term memory, on the other hand, is theoretically limitless, and can store many complicated, often interconnected concepts (Atkinson & Shiffrin, 1968). More experienced physicians are believed to typically have more well-developed illness scripts in their long-term memory; in the case of an internal medicine physician, a more experienced internist likely has a more well-developed illness script for pneumonia, both typical and atypical presentations, than a novice physician or medical student (Bowen, 2006).

Growing directly from an understanding of human memory (and specifically where it intersects the development of knowledge organization, cognitive load theory (CLT) offers a useful explanation for expert performance, weaving together the concepts of memory and knowledge organization. CLT describes three types of cognitive load that are operating within a learner's limited working memory during a task: intrinsic (essential to the task), extrinsic (not essential to the task), and germane (load that is specifically dedicated to the learner attempting to promote long-term learning/memory). If the extrinsic and/or intrinsic loads overburden working memory, then the individual can experience cognitive overload, task performance may suffer and there is unlikely to be any additional bandwidth to encode new schemes into long term memory. Ideally, an educational experience should be designed such that the combination of intrinsic and extrinsic loads leaves enough residual working memory for the germane load necessary to promote long-term learning.

Expert performance in any field is believed to be achieved through deliberate practice, which refers to the process by which a practitioner engages in an activity, under the supervision of a teacher, mentor, or expert who provides feedback and guidance (Ericsson, 1993). The learner makes a deliberate and conscious effort to incorporate that feedback into their practice. The feedback process necessary for deliberate practice can also occur from other feedback method, such as timely feedback of clinical outcomes (e.g. did the resident physician make the correct initial diagnosis for a particular patient that they admitted overnight). In either scenario, deliberate practice requires the practitioner to receive feedback of some kind that they can use to adjust their practice. Ericsson and colleagues have proposed that it takes up to 10 years of deliberate practice to achieve *expert performance* (Ericsson, 2006). In the case of clinical reasoning, a cognitive task that occurs within the physician's brain, promoting deliberate practice is much more difficult than in other, noncognitive tasks. As health professions educators seek to train expert performance in the foundational skill of clinical reasoning, it is crucial to identify opportunities for useful feedback in order to build mechanisms for deliberate practice. Indeed, in the absence of valid and timely feedback mechanisms, the feedback of instructors may have little utility in promoting deliberate practice. If, however, reliable, timely, and valid markers of the underlying cognitive processes were available, this would be of huge value to those seeking to train expert clinical reasoning performance.

Automaticity is also an important part of developing expert performance under cognitive load, illness script, and deliberate practice theories. Indeed, fast (or System 1) thinking allows the physician to free up needed working memory space to learn and perform in a situation. In other words, expert performance is an adaptation facilitated, in part, by sound knowledge organization through experience which also frees up working memory space. For example, while a novice musician must focus great attention on basic functions such as hand positions, an expert has automated (possesses well developed script(s) for) the basic functions and therefore has more cognitive load available for more complex performances and germane load available for devoting toward deliberate practice. Similarly in medicine, expert-performing clinicians are believed to have effectively organized their clinical knowledge into illness scripts that enable them to free up working memory for more in-depth analysis and learning. One way that the interplay of knowledge organization and cognitive load may manifest is using semantic qualifiers (or semantic competence). When a physician readily utilizes semantic qualifiers in this way, they are accessing illness scripts within their long-term memory which allow them to understand and communicate the patient's findings.

Purpose and Research Questions

The purpose of this study is to explore the relationship between semantic competence (and the antithesis of it, semantic dyscompetence) and clinical reasoning performance. We hypothesized that semantic competence is a marker of underlying knowledge organization and semantic dyscompetence is a marker of poor knowledge organization, and that semantic competence ultimately has an impact on the clinical reasoning process and performance.

Research question: What is the relationship between semantic competence and dyscompetence and clinical reasoning performance?

CHAPTER 2: Semantic Competency as a Marker of Clinical Reasoning Performance

This chapter's content was submitted to *Diagnosis* on 27 June 2021.

ABSTRACT

Purpose: This study sought to explore the relationship between semantic competence (or dyscompetence) displayed during think alouds performed by resident and attending physicians and clinical reasoning performance.

Methods: Internal medicine resident physicians and practicing internists participated in "think-alouds" performed after watching videos of typical presentations of common diseases in internal medicine. The think-alouds were evaluated for the presence of semantic competence and dyscompetence and these results were correlated with clinical reasoning performance.

Results: We found that the length of think-aloud was negatively correlated with clinical reasoning performance. Beyond this finding, however, we did not find any other significant correlations between semantic competence or dyscompetence and clinical reasoning performance.

Conclusion: While this study did not produce the previously hypothesized findings of correlation between semantic competence and clinical reasoning performance, we discuss the possible implications and areas of future study regarding the relationship between semantic competency and clinical reasoning performance.

Introduction

At the heart of most health professionals' responsibilities is the task of clinical reasoning, in which the health professional processes clinical and contextual (e.g., situational) information to make patient care decisions (Durning et al, 2013). Clinical reasoning thus encompasses all of the cognitive steps up to and including the arrival at a diagnosis or management plan for a patient (Durning et al, 2010). Clinical reasoning is one of the central pillars of physician performance in any specialty and therefore warrants considerable focus in health care. Indeed, perhaps more than any other subject, medical education should strive to produce clinical reasoning *expert performance* (Durning et al, 2013).

Expert performance in a field is believed to be achieved through *deliberate practice*, in which the individual dedicates deliberate (effortful) practice not just to performing the task but to performing the task with a mindful dedication toward *improving* their performance in that domain, which requires reflection on both the outcomes and the underlying processes (Ericsson, 1993). Achieving expert performance in clinical reasoning arguably requires deliberate practice, and deliberate practice requires that we find ways to assess not just the outcomes of clinical reasoning (i.e., diagnosis and management plans), but also underlying clinical reasoning processes that are occurring within the trainee's mind (i.e., steps proximal to the diagnosis and/or management) (Durning et al, 2013). Given the very nature of clinical reasoning as a thought-based endeavor, it is difficult to design authentic, feasible assessment methods with good reliability and validity evidence for evaluating clinical reasoning processes (Askew et al, 2012; Cambron-Goulet et al, 2019; Lessing et al, 2020; and Daniel et al, 2019). Because of this difficulty in assessment, it is also therefore difficult to design educational experiences to support expert performance in clinical reasoning.

Developing methods to assess clinical reasoning processes requires a detailed understanding of physician thought. Contemporary thinking emphasizes *knowledge organization* as playing a central role in clinical reasoning. Knowledge organization refers to how information is believed to be intertwined (or interdigitated) in memory so that it can be rapidly, effectively, and efficiently recalled and used (Custers et al, 1996). One of the potential windows into one's knowledge organization may be through the words clinicians use when communicating their thought process; from this arises the concept of semantic competence in clinical reasoning. Semantic competence describes the use of abstract, often binary, terms that are used to describe a patient's presentation among physicians (Bordage, 1994; Bordage, 1997). For example, recognizing that a patient's presentation of six hours of wrist swelling, warmth, and tenderness is "acute, monoarticular, arthritis" is an example of semantic competence. Semantic competence is believed to represent an early transformation of knowledge organization and thus a potential target for assessing clinical reasoning processes (Ten Cate and Durning, 2018).

Prior research by Bordage and colleagues has pointed to semantic competence during case presentations as a potential marker of more successful clinical reasoning (meaning that the subject is more likely to arrive at the correct diagnoses in their differential if they manifest semantic competence) (Chang, 1998 and Bordage, 1994). However, additional research into correlating semantic competence with clinical reasoning performance has not shown a clear connection, which may in part be due to the challenges with identifying and achieving consensus with labeling semantic competence as well as a limited number of clinical cases included in the study (see Bordage, 2007, for a discussion of the challenges of studying semantic competence).

Whether or not semantic competence is associated with superior clinical reasoning performance, it is crucial to understand the theoretical underpinning of any possible connection.

In other words, *if* clinicians exhibiting higher degrees of semantic competence have superior clinical reasoning performance, why is that? In order to make this leap, we must first understand the closely related idea of illness script theory, which holds that information about a specific disease is stored in long term memory with varying degrees of accessibility and inter-connectedness based on that clinician's experiences (e.g., a clinician who has seen many patients with a specific disease in many different contexts and presentations may have a better-developed illness script for that disease) (Custers, 2015). Novice clinicians may have a more difficult time accessing the illness scripts in their long term memory due to less development and/or interconnectedness, if such scripts exist at all; this may hold particularly for those who have either not seen a particular illness or who have not yet established well-organized knowledge on the topic in their long term memory (ten Cate and Durning, 2018). Well-developed illness scripts may be associated with arriving at the correct diagnosis more quickly as well as producing superior management plans (Custers, 2015 and Monajemi et al, 2012). There therefore exists the intriguing possibility that semantic competence may serve as a window into the knowledge organization within a clinician's brain, and therefore serve as a marker of the underlying scaffolding necessary to establish expert clinical reasoning performance.

A physician who is able to adeptly express their thoughts about the patient's verbal contribution in a semantically competent manner is likely demonstrating superior knowledge organization. By contrast, physicians with significant prior experience in an area (e.g. medical resident or attending seeing a common disease presentation in their field) who fail to make this articulation, or who do so in a more delayed or even erroneous manner (e.g. semantic dyscompetence), may be demonstrating poor knowledge organization. For example, if a physician is able to successfully express the patient's complaint of "being constantly thirsty and having to urinate a lot" as "polydipsia and polyuria," it is likely that this individual has activated a more organized knowledge network within their long-term memory.

Beyond semantic competence itself, automaticity is also believed to be an important consideration for evaluating the relationship between the words that a physician uses and their underlying mental processes. Dual process theory holds that two distinct thought processes are occuring within the brain, one of which is fast (or system 1) thinking, which is quick, largely subconscious and low effort, and typically used with more knowledge organization (e.g. pattern recognition); and the other of which is slow (or system 2) thinking, which is more deliberate and higher effort. (Croskerry et al. 2014; ten Cate and Durning, 2018). Indeed, physicians displaying more system 1 thinking may perform superiorly in clinical reasoning tasks because they have freed up needed working memory space to learn and perform in the situation. A greater degree of system 1 thinking may manifest through the more automatic expression of (i.e. semantic competency), but the degree of system 1 thinking may also be reflected in the amount of time or number of words they use during their thinking process.

In this study we sought to explore the relationship between semantic competence (and dyscompetence) and clinical reasoning performance. We hypothesized that semantic competence or dyscompetence was a marker of underlying knowledge organization or disorganization respectively, and ultimately would have an impact on clinical reasoning performance. We believed that use of semantic competence would be linked to superior clinical reasoning performance, suspecting that it would be a marker of superior knowledge organization.

More specifically, we made the following hypotheses:

1. The number of dyscompentencies recorded during the think-aloud would be associated with poorer performance on the PEF
2. A physician with no semantic dyscompetencies would perform better on a measure of clinical reasoning performance (a post-encounter form or PEF) .
3. The presence of dyscompetencies would indicate poorer knowledge organization, leading to poorer performance on this PEF.
4. In those physicians who do not display semantic dyscompetence, the number of instances of semantic competence would be associated with higher performance on the PEF.
5. As a marker of knowledge organization and dual process theory, fewer words used by a physician would be associated with superior clinical reasoning performance.

Methods

Study Population and Setting

As part of a larger study conducted at the Uniformed Services University of the Health Sciences (USU), Walter Reed National Military Medical Center (WRNMMC), and Naval Medical Center San Diego (NMCSD) exploring clinical reasoning, active-duty resident and attending physicians in internal medicine, family medicine, and surgery were invited via email to participate. A total of 41 physicians participated, completing two cases each, producing a total of 82 transcripts for analysis. For each physician, one case was a patient presenting with new onset diabetes mellitus and the other presenting with unstable angina. For more details on study design, see both papers from Konopasky and colleagues published in 2020.

Data Collection

After obtaining informed consent, the participants were quasi-randomized (randomization was based on participant schedules) into one of three groups. The participants then watched videotapes of a physician interacting with a patient and then completed a computerized free-text PEF, which asked the participants to provide: 1) what additional history they would obtain from the patient, 2) what additional physical exam findings they would look for, 3) complete problem list, 4) differential diagnosis, 5) leading diagnosis with supporting data, and 6) a treatment plan. Participants had up to thirty minutes to complete the PEF (which was found to be ample time in prior trials). This PEF has reliability and validity evidence for assessing clinical reasoning (Durning, 2011). Items were scored as in prior research where reliability and validity evidence for this instrument were established (Durning, 2012 and McBee, 2017): 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 predetermined scoring key developed by a panel of board certified internists (Durning, 2012 and McBee, 2017), 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.

Immediately following PEF completion, the participants completed a think-aloud protocol (Boren, 2000) as they rewatched the video. The "think-aloud" protocol has been shown in other studies to be an effective way to evaluate clinical reasoning (Durning, 2012; Burbach, 2015; Funkesson, 2007). In the think-aloud procedure the participant verbalizes their thinking as they complete a task with a clear beginning and an end. If the participant does not speak for more than 5 seconds, they are asked to "think-aloud." The think-alouds were transcribed verbatim for analysis.

Data Measurement

In order to assess clinical reasoning performance, we used an abbreviated PEF score, including only the outcome-related items: the questions asking for differential diagnosis, leading diagnosis, and data supporting the leading diagnosis (#6, 7a, and 7b on the PEF).

To assess semantic competence and dyscompetence, a team of coders consisting of one attending physician (study lead), one resident physician, and two medical students reviewed the think-aloud transcripts to identify utterances of semantic competence and dyscompetence. Next, this team of coders collaboratively identified and agreed up on a scoring rubric that was also reviewed and discussed with the entire study team (which included two additional attending physicians), including several sessions collaboratively reviewing transcripts until complete consensus was reached. Our coding rules are shown in Table 1 below.

The scoring rubric provided criteria to standardize scoring for semantic competence and dyscompetence which has been described as a limitation in prior studies (Bordage, 1997). Semantic competence was coded whenever the most appropriate clinical terminology was used by the participant during their think-aloud. Dyscompetence was coded whenever the participant used terminology to describe clinical findings, symptoms, or diagnoses that were not considered the most clinically-appropriate terminology, as discussed above.

The coding team reached consensus on coding rules/rubric (Table 1) after reviewing approximately a quarter of the transcripts (n = 20), at which point the remainder of the transcripts were assigned to both the study lead and one additional coder. After coding independently, the study lead met with the coders and came to complete consensus for the coding of each transcript. Disagreement, when present, was resolved by referencing the previously agreed-upon coding rules and analyzing the specific utterance. The initial 20 transcripts were also re-coded independently based on the final coding rubric and complete consensus was achieved, as with the other transcripts.

Data Analysis

After all of the transcripts were coded ($n = 82$), the total number of instances of semantic competence and dyscompetence were recorded for each transcript. Total word count of each think-aloud was also recorded.

Descriptive analysis was performed on all variables to understand the patterns of competence and dyscompetence. To address our hypotheses, we ran correlational analyses to evaluate the relationship between word count, the various coded totals of semantic competence, semantic dyscompetence, and clinical reasoning performance, measured by the participant's score on the PEF.

Results

There were 41 participants in the study, each of whom reviewed two cases and participated in two think-aloud sessions, resulting in a total of 82 transcripts. The demographics of the participants are detailed in Table 2 below.

The coders evaluated the transcripts for instances of semantic competence or dyscompetence as defined by the rules in Table 1. Descriptive statistics of this evaluation are listed in Table 3.

In an effort to evaluate our hypotheses that quantity of speech and the degree of semantic competence and dyscompetence would be associated with clinical reasoning performance, we ran Pearson correlations (*n* = 82) between PEF score and 1) word count, 2) total competencies, 3) total dyscompetencies, and 3)the total dyscompetencies (with 2-tailed significance), as shown in Table 4.

The main finding was a statistically significant but small negative correlation between think-aloud word count and the participant's clinical reasoning performance (p = .008). A small and weak negative correlation was also seen between the total competency count and the clinical reasoning performance, which is thought to be related to the small/weak positive correlation that was seen between word count and both competencies and dyscompetencies (in other words, participants who spoke more, used both more competencies and dyscompetencies).

In order to evaluate our hypothesis that presence of dyscompetencies may indicate poorer knowledge organization, a comparison of mean PEF score was performed between those with zero dyscompetences (*n* = 26) and those with greater than zero dyscompetencies (*n* = 56), which yielded no statistically significant difference (see table 5, p = .242)

To evaluate if, among those who did not display semantic dyscompetence, the degree of semantic competence was associated with superior clinical reasoning performance, we ran a Pearson correlation within this cohort (*n* = 26) between abbreviated PEF score and word count and total competencies (see Table 6). No statistically significant relationships were found.

Discussion

In order to assess the potential link between semantic competence and clinical reasoning performance, we evaluated the transcripts of think-alouds performed by physicians after watching standardized clinical case videos for semantic competence/dyscompetence. Using a post-encounter form with reliability and validity evidence, we attempted to correlate this semantic competence or dyscompetence with clinical reasoning outcomes. Based on illness script and dual process theory, we sought to explore the relationship between semantic competence, semantic dyscompetence, and clinical reasoning performance, forming several hypotheses around how these markers of knowledge organization (clinical reasoning process measures) would be related to clinical reasoning performance (outcomes). (Bordage, 1994; Bordage, 1997; Bordage, 2007; Ten Cate and Durning, 2018).

First, we hypothesized that both the presence (e.g., making an utterance) and the degree (e.g., number of such utterances) of semantic dyscompetence would be associated with poorer clinical reasoning performance. This was not demonstrated in our results. This may have reflected the success of the think-aloud process whereby a participant may state what comes to mind without analyzing their thinking, or possibly that participants interpreted the think-aloud prompt as directing them to use layman's terms, since the researcher asking the questions was not a physician. A future study could explore more formal presentations where the participant would be expected to prepare their presentation. The presence of semantic dyscompetence in this setting may be a more sensitive marker of potential knowledge disorganization. In other words, the resident who says "peeing a lot" in lieu of "polyuria" during a patient presentation on rounds may be demonstrating a different phenomenon than the physician who states "peeing a lot" during a think-aloud exercise.

Our other hypothesis from illness script theory regarding the relationship between semantic competence and clinical reasoning performance was also not confirmed. As was the case above, the effect of the think-aloud protocol on participant thinking was likely at work as in semantic dyscompetence. Additionally, the cases that were used in this study were deemed to be straightforward presentations for the experience level of the physicians participating in the study. Thus, for both semantic competencies and dyscompetencies, these relatively less complex cases may not have been sensitive of superior or poor clinical reasoning performance. Future work with more complex cases could reveal a relationship between semantic competencies and dyscompetencies and clinical reasoning performance.

Finally, we also hypothesized that a longer word count would be associated with poorer clinical reasoning performance, and this was verified by our data, which demonstrated a negative correlation between word count and abbreviated PEF score. This is theoretically supported by dual process theory, in that physicians who have less well-developed illness scripts may have "meandered" more in their think-alouds, whereas users with more well developed scripts may have been more succinct and accurate. (Croskerry et al. 2014; ten Cate and Durning, 2018).

Our study had several limitations. We only gave participants two clinical cases each and while the study contained a moderate number of participants for such a study, the findings may have been different with a larger sample. Also, while our PEF has reliability and validity evidence in this context, it is just one measure of clinical reasoning outcomes.

While our study did not confirm many of our hypotheses, it did demonstrate a slight negative correlation between think-aloud length and clinical reasoning performance. We believe that there is still potential educational value in looking into the link between semantic competence, as a marker of knowledge organization, and clinical reasoning performance. Medical and other health professional educators need ways to assess clinical reasoning processes as well as outcomes to enhance patient care. As we strive to promote expert performance in clinical reasoning amongst our trainees, we need to continue to seek markers of the clinical reasoning process and underlying knowledge organization. Despite the largely negative results of our study, there is still a strong theoretical framework for further study of the link between semantic competence, knowledge organization, and clinical reasoning.

Conclusion

We explored the association between a proxy of knowledge organization (semantic competencies and dyscompetencies) and clinical reasoning performance. The length of the think-aloud was negatively correlated with clinical reasoning performance, supporting dual process theory; however the remainder of our hypotheses about the relationship between semantic competence/dyscompetence and clinical reasoning performance were not supported. Semantic competence, however, may yet be a potentially useful marker for knowledge organization, despite the lack of statistically significant results in our analysis. This study prompts many follow-up questions, most specifically whether the nature of semantic competence/dyscompetence would change in the setting of formal case presentations or work rounds as opposed to the think-aloud format that we evaluated.

Table 1: Coding Rules	
Competency Rules	
1.	Semantic competence was coded on the lexical level (i.e., can have more than one semantic competence code within a sentence).
2.	In order to be considered an instance of semantic competence, an utterance must be the most appropriate medical term used in an appropriate clinical context. Specific examples include: a. Fatigue (instead of tired/tiredness) b. Exertional (instead of "when doing things" or "with activity")
3.	The appropriate term for a disease process must be used to be considered semantic competence: a. "Gastroesophageal reflux disease" or "GERD" would be considered competence. "Reflux" or "heartburn" would be considered dyscompetent.
4.	All appropriately-used descriptors (e.g. acute vs chronic) for signs/symptoms/findings were coded separately as semantic competence utterances. For example: "acute progressive dyspnea" would be coded as three separate instances of semantic competency.
5.	Semantic competence was also coded if used properly in the negative sense (e.g., "The patient is not tachycardic").

Dyscompetency Rules	
6.	Semantic dyscompetence was coded when the participant used lay terminology when more appropriate medical terminology existed (e.g. "the patient is drinking a lot" in lieu of "polydipsia")
7.	Dyscompetence was also specifically documented when the participant used semantically competent language but *subsequently* used the analogous semantically dyscompentent terminology (e.g. "the patient has polyuria. They are peeing a lot.")
Exclusion Rule	
8.	Anything that was repeated (i.e., parroting patient's words) without demonstrating a transformation was not included in analysis (e.g., if the participant used lay terminology, but they were repeating the words from the patient in the video, such as "he says he's really thirsty," this was not coded)

Table 2: Demographics of Study Participants	
Average age	35
Average years of practice (since medical school graduation)	7.2
Percent female	27%

Table 3: Descriptive Statistics on Think-Aloud Coding		
Variable	Average (standard deviation)	Range (minimum-maximum)
Abbreviated PEF score	2.00 (0.525)	0.67-2.76
Word Count	648 (419)	190-2094
Competencies: Total	11.1 (5.33)	1-30
Dyscompetencies: Total	1.26 (1.39)	0-7

Table 4: Correlations						
Variable	M	SD	1	2	3	4
1. Abbreviated PEF score	2.00	.525	-			
2. Word Count	648	419	-.292**	-		
3. Competencies: Total	11.1	5.33	-.221*	.274*	-	
4. Dyscompetencies: Total	1.26	1.39	.029	.224*	.036	-
*p < .05. **p < .01.						

Table 5: Comparison of Abbreviated PEF between Dyscompetent and Non-Dyscompetent Groups				
Groupings:	N	Mean Abbreviated PEF	Std Dev	Std. Error Mean
Dyscompetencies > 0	56	2.07	.473	.063
Dyscompetencies = 0	26	1.85	.608	.119

Table 6: Correlations Within the Transcripts that Demonstrated Zero dyscometencies (N=26)

Variable	M	SD	1	2	3
1. Abbreviated PEF score	1.91	.501	-		
2. Word Count	663	463	-.170	-	
3. Competencies: Total	10.6	5.53	.084	.129	-
*p < .05. **p < .01.					

CHAPTER 3: Discussion

Despite the general dearth of statistically significant results found in this research project, elucidating the mechanisms underlying clinical reasoning remains crucial to unlocking our potential as educators and clinicians to pursue expert clinical reasoning performance in health professions trainees. As discussed above, deliberate practice necessitates that we continue to search for markers of the processes which underlie clinical reasoning that have validity evidence. The evidence that shorter word count is associated with superior clinical reasoning performance is consistent with our theoretical framework, supported by the concepts of knowledge organization and automaticity. The theoretical underpinnings of the remainder of the hypotheses in this thesis remain sound despite the lack of statistically significant findings to support them. The concepts of knowledge organization, illness script theory, and the relationship with semantic competence remain promising areas of inquiry despite the results.

Military Relevance

Establishing and maintaining expert performance in the civilian world is a challenging endeavor, fraught with myriad challenges described above. These challenges are compounded further by the unique military medical environment. Military physicians strive to achieve clinical mastery in their trade through the traditional realms of undergraduate medical education and graduate medical education. Military healthcare professionals (at all levels from medic to attending physician) are also expected to be ready to perform at an expert level in a much broader spectrum of tasks than their civilian counterparts, often in tasks that may be rarely performed (e.g. a General Medical Officer in a battlefield medicine or deployed public health scenario). While the education of all healthcare professionals could be greatly enhanced by improved tools for assessing and developing expert clinical reasoning performance, this need is even more pressing within the military. Just as the implications of continued research in this arena may help drive more thoughtful instructional design in UME and GME, deeper insights into how learners may display both knowledge organization and at times cognitive overload would also have valuable implications for educating military healthcare professionals, particularly in tasks for which frequent, deliberate practice is more difficult to obtain, such as deployed or operational tasks.

Limitations

There are various possible reasons for these findings. First and foremost, the think aloud format may not be the best for evaluating the significance of semantically competent or dyscompetent language being used by clinicians. The format encourages physicians to describe what is in their minds without filtering or analyzing their thought processes. It may be that semantic competence demonstrated in a more formal setting, such as in case presentations or with colleagues, is a more reliable indicator of knowledge organization and therefore of clinical reasoning performance. Also, the cases that were used for this study were relatively straight-forward presentations of common internal medicine disease processes. It is possible that more challenging cases may yield different results.

Future Research and Practice Implications

Semantic competence, as it was defined in this study and within the context of these cases, may not have proven to be a significant marker of clinical reasoning performance, but the theoretical frameworks used through this thesis work serves as a promising foundation for continuing to search for markers of clinical reasoning performance. The research team was surprised by the findings that semantic dyscompetence was not statistically significantly associated with clinical reasoning performance. It is theoretically reasonable that the think-aloud format is a cause of this phenomenon. In other words, since participants were explicitly told *not* to filter their thoughts during the think aloud, they may have suppressed semantic transformations that they would have made if they had not been given that guidance.

The major lines of research that I would propose would be to apply similar research principles to formal case presentations and cases involving more challenging clinical cases. As mentioned above, the think aloud format itself may have significant limitations for evaluating the validity of semantic competence as a marker of knowledge organization. Furthermore, the think aloud is not currently a part of routine clinical practice, whereas the case presentation is already central to clinical medical education, in both the undergraduate and graduate medical education environment.

As discussed above, the cases used in this study were relatively straightforward presentations of common disease processes. If a similar research approach was applied to more difficult cases, it would be interesting to see if more significant results would be obtained. It is likely that more challenging cases would be more likely to put the participants closer to the edge of their cognitive load, possibly forcing them into a greater degree of semantic dyscompetence. Also, more difficult cases would likely make for a broader distribution of clinical reasoning performance outcomes, which could make it easier to find statistically significant relationships between clinical reasoning performance and the semantic factors, for both studies on think alouds and on case presentations.

One of the other areas which was not addressed in this study is the context specificity of think alouds. Another possible avenue of future research would be to look at the relationship between contextual factors, semantic competence, and clinical reasoning performance. Specifically, this research could evaluate the theoretical link between contextual factors on cognitive load. Given that clinical reasoning performance may be influenced by contextual factors, could semantic competence in these cases be a marker of cognitive load, and whether or not a clinician is operating in an environment where they are close to their cognitive load limit or not.

Through some of the research opportunities described above, if semantic competence/dyscompetence is associated with superior knowledge organization or outcomes, then further research may allow for development of assessment tools to enable more complex instructional design while building toward expert clinical reasoning performance. This insight would be helpful in designing instruction and assessment designed to promote learning in the field of clinical reasoning. For example, if semantic dyscompetence (perhaps in a formal setting such as case presentation instead of think aloud) is associated with inferior clinical reasoning performance, it would provide a valid educational foundation for instructors to identify clinicians who may need additional instruction in clinical reasoning in general or in specific cases. Without such evidence, or if strong evidence of no correlation were to be discovered, then it would be reasonable to focus educational efforts on other areas of clinical reasoning performance and correcting presentation competence could be viewed more as a matter of "professional polish" than a matter of foundational competence. While this may seem like a minor distinction, case presentation is often used by faculty currently a major subjective (i.e. nonstandardized "impression-based") indicator of clinical competence that is difficult to quantify.

REFERENCES

Askew K, Manthey D, Mahler S. Clinical reasoning: are we testing what are teaching?. Med Educ 2012; 46:540–542.

Atkinson RC and Shiffrin RM. (1968). Human memory: A proposed system and its control processes. In K. W. Spence & J. T. Spence, The psychology of learning and motivation: II. Academic Press

Battista A, Konopasky A, Ramani D, et al. Clinical reasoning in the primary care setting: two scenario-based simulations for residents and attendings. Med Ed Portal 2018; 14: 10773.

Bordage G. (1994). Elaborated knowledge: a key to successful diagnostic thinking. Academic Medicine, 69, 883–885.

Bordage G. Prototypes and semantic qualifiers: from past to present. Medical Education 2007; 41: 1117-1121

Boren T. and Ramey J. "Thinking Aloud: Reconciling Theory and Practice" IEEE TRANSACTIONS ON PROFESSIONAL COMMUNICATION, VOL. 43, NO. 3, SEPTEMBER 2000 261

Bowen, J. (2006). Educational Strategies to promote clinical diagnostic reasoning. The. New England Journal of Medicine, 355(21), 2217-2225.

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.

Cambron-Goulet E, Dumas J, Bergeron E, Bergeron L, St-Onge C. Guidelines for Creating Written Clinical Reasoning Exams: Insight from a Delphi Study, Health Professions Education, Volume 5, Issue 3, 2019, Pages 237-247.

Chang RW, Bordage G, Connell KJ, (1998) "Cognition, Confidence, and Clinical Skills." Academic Medicine, 73, S109-S111.

Croskerry P. A universal model of diagnostic reasoning. Acad Med. 2009;84(8):1022-1028.

Custers E, Regehr G, and Norman G. (1996). Mental representations of medical diagnostic knowledge: A review. Academic Medicine, 71(10), S55–S61.

Daniel, Michelle MD, MHPE; Rencic, Joseph MD; Durning, Steven J. MD, PhD; Holmboe, Eric MD; Santen, Sally A. MD, PhD; Lang, Valerie MD, MHPE; Ratcliffe, Temple MD; Gordon, David MD; Heist, Brian MD, MSc; Lubarsky, Stuart MD, MHPE; Estrada, Carlos A. MD, MS; Ballard, Tiffany MD; Artino, Anthony R. Jr PhD; Sergio Da Silva, Ana PhD; Cleary, Timothy PhD; Stojan, Jennifer MD, MHPE; Gruppen, Larry D. PhD Clinical Reasoning Assessment Methods: A Scoping Review and Practical Guidance, Academic Medicine: June 2019 - Volume 94 - Issue 6 - p 902-912 doi: 10.1097/ACM.0000000000002618

Durning, SJ, Artino AR, Pangaro LN, van der Vleuten C, and Schuwirth L. (2010a). Redefining context in the clinical encounter: Implications for research and training in medical education. Academic Medicine, 85, 894–901

Durning, SJ, Artino AR., Pangaro L., van der Vleuten C, Schuwirth L. (2011) Context and clinical reasoning: understanding the perspective of the expert's voice. Medical Education; 45: 927–938

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 Health Sci Educ 2012; 17(1): 65–79.

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. Med Teach. 2012;34(1):30–7.

Durning, S.J., Artino, A.R., Schuwirth, L., Cees van der Vleuten, C. (2013). Clarifying Assumptions to Enhance Our Understanding and Assessment of Clinical Reasoning

Elstein AS, Shulman LS, Sprafka SA. Medical problem solving an analysis of clinical reasoning. 1978.

Ericsson KA, Krampe RT, and Tesch-Romer C. The Role of Deliberate Practice in the Acquisition of Expert Performance.Psychological Review 1993, Vol. 100. No. 3, 363-406

Ericsson KA. Protocol analysis and expert thought: Concurrent verbalizations of thinking during experts' performance on representative tasks. In: Ericsson KA, Charness N, Feltovich PJ, Hoffman RR, eds. The Cambridge Handbook of Expertise and Expert Performance. New York, NY: Cambridge University Press; 2006:223–241.

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.

Heneghan C, Glasziou P, Thompson M, et al. Diagnostic strategies used in primary care. Bmj. 2009;338(apr20_1):b946-b946.

Konopasky A., Ramani D, Ohmer M, Battista A, Artino AR. McBee E, Ratcliffe T, and Durning SJ. "It Totally Possibly Could Be: How a Group of Military Physicians Reflect on Their Clinical Reasoning in the Presence of Contextual Factors." MILITARY MEDICINE, 185, S1:575, 2020

Konopasky A, Artino AR, Battista A, Ohmer M, Hemmer PA, Torre D, Ramani D, van Merrienboer J, Teunissen PW, McBee E, Ratcliffe T, Durning SJ. Understanding context specificity: the effect of contextual factors on clinical reasoning. Diagnosis (Berlin, Germany), 2020-09, Vol.7 (3), p.257-264

Lessing, JN, Rendón, P, Durning, SJ, Roesch, JJ. Approaches to Clinical Reasoning Assessment, Academic Medicine: August 2020 - Volume 95 - Issue 8 - p 1285

Mamede S, van Gog T, Sampaio AM, de Faria RMD, Maria JP, Schmidt HG. How Can Students' Diagnostic Competence Benefit Most From Practice With Clinical Cases? The Effects of Structured Reflection on Future Diagnosis of the Same and Novel Diseases, Academic Medicine: January 2014 - Volume 89 - Issue 1 - p 121-127 doi: 10.1097/ACM.0000000000000076

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.

Monajemi, A. , Schmidt, H. G. & Rikers, R. M. (2012). Assessing Patient Management Plans of Doctors and Medical Students: An Illness Script Perspective. Journal of Continuing Education in the Health Professions, 32(1), 4–9. doi: 10.1002/chp.21118.

Norcini J, Anderson B, Bollela V, et al. Criteria for good assessment: consensus statement and recommendations from the Ottawa 2010 Conference. Med Teach. 2011;33(3):206–214.

Sam AH, Hameed S, Harris J, et al. Validity of very short answer versus single best answer questions for undergraduate assessment. BMC Med Educ. 2016;16(1):266.

Scouller K. The influence of assessment method on students‘ learning approaches: multiple choice question examination versus assignment essay. High Educ. 1998;35(4):453–472

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.

Young M, Thomas A, Lubarsky S, et al. Drawing boundaries: the difficulty in defining clinical reasoning. Acad Med. 2018;93(7):990-995.