Some assembly required: tracing the interpretative work of Clinical Competency Committees

Rachael Pack,[1] Lorelei Lingard,[1,2,3] Christopher J Watling,[1,4] Saad Chahine[1,2,3] & Sayra M Cristancho[1,2,5]

OBJECTIVES This qualitative study describes the social processes of evidence interpretation employed by Clinical Competency Committees (CCCs), explicating how they interpret, grapple with and weigh assessment data.

METHODS Over 8 months, two researchers observed 10 CCC meetings across four postgraduate programmes at a Canadian medical school, spanning over 25 hours and 100 individual decisions. After each CCC meeting, a semi-structured interview was conducted with one member. Following constructivist grounded theory methodology, data collection and inductive analysis were conducted iteratively.

RESULTS Members of the CCCs held an assumption that they would be presented with high-quality assessment data that would enable them to make systematic and transparent decisions. This assumption was frequently challenged by the discovery of what we have termed 'problematic evidence' (evidence that CCC members struggled to meaningful interpret) within the catalogue of learner data. When CCCs were confronted with 'problematic evidence', they engaged in lengthy, effortful discussions aided by contextual data in order to make meaning of the evidence in question. This process of effortful discussion enabled CCCs to arrive at progression decisions that were informed by, rather than ignored, problematic evidence.

CONCLUSIONS Small groups involved in the review of trainee assessment data should be prepared to encounter evidence that is uncertain, absent, incomplete, or otherwise difficult to interpret, and should openly discuss strategies for addressing these challenges. The answer to the problem of effortful processes of data interpretation and problematic evidence is not as simple as generating more data with strong psychometric properties. Rather, it involves grappling with the discrepancies between our interpretive frameworks and the inescapably subjective nature of assessment data and judgement.

Medical Education 2019: 53: 723–734
doi: 10.1111/medu.13884

[1]Centre for Education Research & Innovation, Schulich School of Medicine and Dentistry, Western University, London, Ontario, Canada
[2]Faculty of Education, Western University, London, Ontario, Canada
[3]Department of Medicine, Schulich School of Medicine and Dentistry, Western University, London, Ontario, Canada
[4]Department of Clinical Neurological Sciences, Schulich School of Medicine and Dentistry, Western University, London, Ontario, Canada
[5]Department of Surgery, Schulich School of Medicine and Dentistry, Western University, London, Ontario, Canada

Correspondence: Rachael Pack, Centre for Education Research and Innovation, Schulich School of Medicine and Dentistry, Western University, Health Sciences Addition, Room 110, London, Ontario N6A 5C1, Canada. E-mail: rachael.pack@schulich.uwo.ca

INTRODUCTION

The turn towards competency-based medical education (CBME) necessitates the implementation of new, multifaceted programmes of assessment that generate a wide range of data about learners' progression towards competency.[1–4] The large volume of diverse data produced, in turn, requires new processes of assembly and analysis. In response to this need, postgraduate medical programmes in Canada and the USA have integrated new group decision-making bodies, referred to as Clinical Competency Committees (CCCs), into their programmes of assessment.[5] These small groups operate as centralised review bodies that conduct in-depth reviews of diverse assessment data collected about trainee performance and use collective processes to render decisions about progression.[6,7]

This type of small group decision making is no small task. Committees must sift through large amounts of quantitative and qualitative performance data, interpret their meanings, and create an intelligible composite that reliably communicates the trainee's competencies.[5,8,9] This process relies on the expertise and judgement of committee members and their ability to interpret performance evidence and transform it into a decision about trainee competence. The process of synthesising heterogeneous assessment data is fraught with challenges[10–12] and the need for efficient, creative strategies to support the complex task of data aggregation has been communicated in the literature.[13,14]

Given the high-stakes nature of progression decisions, the decision-making processes, data review procedures and the operating practices of CCCs have been subject to recent scholarly attention. Concerns around the fairness and transparency of progression decisions have been at the heart of this conversation.[8,11,15] The current lack of systematic procedures for evidentiary review, weighting and interpretation[16,17] and the heterogeneity of practices within committees have been problematised.[9,10] In response, a variety of strategies have been proposed to standardise CCCs, including best operating practices,[14,16,17] faculty training,[10,15] data visualisation software[13,18] and data analysis programs.[8]

A central thread of the scholarly discussion on CCCs has been focused on the quality of evidence produced for CCC review and has drawn attention to the issue of poor or variable data quality.[10] Underpinning this conversation is an assumption that performance data generated for review can be objective – or that they can reliability communicate the 'truth' about workplace performance and competencies. From our perspective as social constructivists rooted in post-modernism, this assumption belies the possibility that 'objective' evidence may not exist[19] and obscures the socially situated processes involved in evidence generation and interpretation.

To date, there is no discussion in the literature about how members of CCCs engage in the socially situated processes of evidence interpretation. We suggest that how CCC members come to derive meaning from and make sense of performance evidence is deserving of critical attention, beyond the question of data quality. Thus, the purpose of this constructivist grounded theory study is to explore the socially constructed and culturally specific interpretive frameworks that CCC members draw on in their appraisal of evidence.

METHODS

Setting and context

Our research was conducted at one Canadian medical school. At this medical school, all programmes have been mandated to form a CCC, regardless of the anticipated timing for each programme's nationwide implementation of a CBME curriculum. In Canada, the Royal College of Physicians and Surgeons has provided postgraduate programmes with a list of recommendations regarding the processes and procedures of CCCs[20] but has granted autonomy and flexibility in implementation. As a result, each of the CCCs at this institution developed their own terms of reference to suit the specific needs, context and structure of their training programme. Four postgraduate programmes, diverse with respect to specialty and size, were invited to participate. Two of the programmes had already adopted CBME, whereas the remaining two programmes were preparing for the educational transition and continued to use more traditional assessment tools (i.e. in-training evaluation reports [ITERs]). The programmes of assessment utilised by each participating CCC are reported in Table 1.

Table 1 Summary of the assessment strategies reviewed by Clinical Competency Committees

	Entrustable professional activities (EPAs) with numeric rating and narratives	End-of-rotation evaluation (numeric rating and narrative)	Free-form narratives from faculty members	Free-form narratives from peers	Free-form narratives from academic advisors	Standardised written examinations	Oral examinations	Standardised clinical examinations (OSCE)	Trainee-maintained logbooks	Clinical evaluations (numeric rating)
CCC1	✓		✓		✓	✓			✓	
CCC2		✓	✓		✓	✓	✓			✓
CCC3		✓	✓		✓	✓	✓			
CCC4	✓		✓	✓		✓		✓	✓	✓

The CCCs included in this study had a common membership structure, consisting of a committee chair, the postgraduate programme director, the programme administrator and at least four faculty members involved in postgraduate training. Participating CCCs met between four and 12 times per year.

Given postgraduate programmes' high degree of autonomy, the operating practices and procedures of each CCC in this study were unique, and thus identifiable. To provide sufficient context for the interpretation of our findings while preserving the anonymity of our participants we present an aggregate description of their procedures.

Each resident was assigned one or two faculty reviewers, the assignments of which typically rotated each meeting. Although the type of data reviewed varied by programme (see Table 1), it generally included numeric ratings of trainee performance and narrative feedback provided by faculty members, academic advisors and, in some cases, peers. Presentations of trainee data were delivered by the reviewers in a standardised format (unique to each programme) in which the member provided a synthesis of each type of assessment data (i.e. examination score, narrative comments, entrustable professional activity assessments). Following this presentation, the CCC discussed the data and, if deemed necessary, collectively reviewed specific pieces of evidence (i.e. a single end-of-rotation report or narrative comment). Depending on the nature of the decision at hand, the committee would either take a formal vote (in the case of progression decisions) or determine through informal consensus (without voting procedures) that the trainee was, or was not, progressing as expected.

Data collection and analysis

Study data were collected between September 2017 and July 2018. In 25 hours of observation, two members of the research team (R.P. and S.C.) collected field note data regarding 10 CCC meetings, including approximately 100 individual progression decisions. CCC meetings ranged between 90 minutes and 4 hours in length and four participating CCCs were observed between two and four times. Both researchers took detailed field notes during the meetings to capture their observations about how the processes of data interpretation unfolded and the character of the discussion that informed each individual decision, including points of agreement or disagreement, the language used to describe specific types of data and direct quotations of CCC members. In alignment with the principles of grounded theory, the structure and content of the field notes were refined throughout the study to reflect and sample for our developing analytic insights.[21]

After each CCC meeting, one CCC member participated in a semi-structured (n = 10) interview. Interview participants were purposefully selected according to their role on the CCC, faculty position and career stage, in order to capture a broad range of perspectives. Interview participants included committee chairs (n = 2), programme directors (n = 2), programme-specific CBME experts (n = 2) and general committee members (n = 4). The purpose of the interview was to gain additional insight into the unique, context-specific data interpretation processes of the CCC from the perspective of individual members. Although the interviews were semi-structured, the nature and content of the questions we asked each participant were responsive to the observational data; therefore,

each interview had its own unique interview guide. For example, if we observed a discussion at the preceding CCC meeting where members disagreed over the interpretation of a particular piece of evidence, the participant would be asked to provide their perspective on that discussion. This practice of data-responsive interviewing aligns with the exploratory nature of qualitative research techniques[22] and enables theoretical sampling for emerging insights.

Interviews ranged between 30 and 90 minutes in length, were audiorecorded and transcribed verbatim. Sampling and data collection were suspended once theoretical sufficiency[23] was reached and the refined categories were robust enough to capture the variations present within the data. This study received ethical approval from the Western University institutional review board.

Given the sensitivity of trainee performance data, this study was designed in negotiation with our clinical collaborators (key informants on each CCC) and was intended to ensure the acceptability of both our methods and ethical procedures to safeguard identifying information about trainees and high-stakes progression decisions. No identifying data about trainees were collected; the focus of our research was specifically concerned with the social processes of data interpretation that occurred during decision-making conversations, rather than the decisions themselves.

This study was conducted using constructivist grounded theory (CGT),[24] a methodology that is well suited for the exploration of social processes (in this case, the social processes of small group decision making) and the socially situated nature of CCCs.[25] Acknowledging the creative and vital role of the researchers and their unique perspectives in the production of grounded theory,[26] we assembled an interdisciplinary research team that included PhD researchers with expertise in communications, social sciences, assessment and group dynamics, as well as clinician educators with frontline experience implementing CBD curriculum and convening CCCs. We met frequently throughout the data collection and analysis processes to discuss our emerging findings and collaboratively generate analytic insights.

Consistent with the iterative processes of CGT,[26] field notes and interview transcripts were analysed using constant comparative techniques across the data collection period to facilitate the identification and refinement of recurring themes and the relationship between identified concepts. Data collection and analysis occurred simultaneously, enabling the results of our ongoing data analysis to inform subsequent data collection. NVIVO VERSION 11.4 (QSR International, Melbourne, Victoria, Australia) software was used to support data analysis.

In the initial stage of data collection, one researcher (RP) closely read the interview transcripts and field notes and applied open coding techniques to identify recurring patterns. Following each instance of data acquisition, emerging findings were presented, discussed and refined at an analysis meeting with all members of the research team (SC, SMC, CW and LL). A set of initial codes were collaboratively developed and applied to a set of field notes and transcripts by RP and SMC to determine fit. The results of open coding were reviewed and discussed by the broader research team and opportunities to theoretically sample[24] emerging insights were identified. For instance, as we observed committees struggle to make sense of evidence they perceived to be problematic we amended our interview strategy to explore this unanticipated phenomenon, selected interview participants who were substantively involved in discussions around problematic evidence and used observed instances of problematic evidence as interview probes.

Following open coding, two researchers (RP and SMC) worked in tandem to consolidate the initial codes into a coding structure organised around recurring themes and conceptual threads. To test the salience of the coding structure, RP and SMC individually applied the structure to four sets of field notes and corresponding interviews, identifying examples that confirmed or challenged the codes. They met throughout this process to refine the coding structure until it appeared to be salient and stable. The coding structure was presented to the rest of the research team (LL, SC and CW) along with excerpts from remaining uncoded data. Team members applied the coding structure to the data and provided feedback about the salience of the codes to collectively refine existing boundaries and properties of our categories and to articulate the relationships between them. The final coding structure was applied to the entire dataset by RP.

In subsequent meetings, the research team, informed by our coding structure, directed our

analytic attention towards articulating a conceptual, explanatory model that robustly described the social processes and phenomena we observed. This theoretical model is included as Fig. 1 in the results section of this paper.

To further explore the resonance of the theoretical model we derived from our data analysis, we presented the model to three clinical faculty members who did not participate in the study but who were engaged in similar CCC work at our own and two other institutions. In these meetings, synthesised data were presented individually to clinical faculty members with experience administering or participating in CCCs, who were asked to describe whether or not the results resonated with their own experiences.

RESULTS

We observed two patterns of data interpretation, which we characterised as 'effortless' and 'effortful'. Most decisions appeared straightforward in that the trainee data came together to tell a cohesive, intelligible story. In these instances, data interpretation appeared to be effortless; the committee members quickly reached consensus and little discussion or debate of the evidence was required. However, we observed other instances where evidence was not easy to make sense of and the committee had to grapple with various pieces of data to determine what counted as evidence and what was actionable. These moments appeared to arise when what we have termed 'problematic evidence' was introduced. In what follows we briefly describe the pattern of effortless data interpretation, before turning our attention to the focus of this paper: more effortful moments of data interpretation and the phenomenon of problematic evidence.

Members of the CCCs held an assumption that the data collected on each learner would be of high quality and would thus enable them to make transparent decisions using standardised processes. Informed by this assumption, participants frequently described their group's data interpretation process as one driven by quantitative data to the exclusion of subjective information about the learner. Outwardly, the committees strove to adhere to a 'follow the numbers' approach whereby composite metrics were transformed into decisions. For example, one committee interpreted scores of four or above on a five-point entrustability scale as successfully completed assessments, or 'clear passes.' Thus, if trainees met the required number of successful observations for any particular EPA, the EPA was marked as 'achieved' without any subsequent discussion or dissection of the data.

When CCC's encountered evidence that they perceived to be straightforward it enabled a process of data interpretation that appeared to be almost effortless in that it required little debate or evidentiary grappling. The following field note excerpt illustrates how effortless data interpretation unfolds in the context of a resident review:

> Committee member 3 reports that resident 10's numbers are on par with the other trainees in his cohort; he has already achieved three EPAs. The trainee's academic advisor submitted written feedback to the committee indicating that resident 10 is "on track" with his assessments, specifically he is requesting the required number of observations and has not received any complaints or negative feedback from faculty. Committee member 3 reports that the narrative comments that she reviewed were all positive but doesn't provide any specific examples. Resident 10 received 19 daily performance evaluations (numeric rating scale) with an average score of 3.6. This score is consistent with the average for the trainee's cohort. The committee chair comments that based on the review thus far, it does not appear that there are any red flags in the data. No further discussion about resident 10 occurs. (Field note 17)

In the above presentation of one trainee's data, we can see how the evidence presented by committee member 3 was interpreted by the committee members as straightforward and thus enabled them to 'follow the numbers' to reach a determination that the resident was progressing as expected without a thorough dissection of the data.

In contrast with the above systematic process that was often marked by an *absence* of debate, we observed numerous instances where committees engaged in a more 'effortful' process of data interrogation and dissection. This departure was necessitated by what we have termed 'problematic evidence': evidence that the committee perceived to be of questionable quality, validity or reliability. Pervasive rather than aberrant, problematic evidence arose at each CCC meeting and was found in the assessment data of both residents who were

progressing as expected and those who were not. Although the phenomenon of effortless data interpretation is certainly worthy of in-depth examination it is not the focus of this paper; we are concerned primarily with problematic evidence. Specifically, in this paper, we examine moments of effortful data interpretation to explicate the interplay between the socio-culturally structured mandates that shape how data and objectivity were conceptualised by the members of the committee (i.e. interpretive frameworks) and socially situated practices of collective data interpretation (through which meaning is derived from data) to illuminate the tensions that arise.

In what follows, we describe the phenomenon of problematic evidence and the social processes of data interpretation, analysis and assembly it generates. To preserve the rich and situated detail of our data we present our results in the form of a story. Although the effortful discussion we present is derived from a real case we observed, all details specific to the trainee and postgraduate programme have been anonymised. Excerpts from participant interviews and field notes have been woven into this representative story to maintain the integrity of our results. Adopting this strategy enables us to fulfill our ethical commitments to our participants, while authentically representing the phenomenon of problematic evidence.

The story

The committee member responsible for presenting the learner's assessment data begins by describing Alex, a junior trainee, as a 'hard worker' who is well liked by the faculty and a 'leader of the pack'. The presenter informs the group that the learner has achieved the highest number of successfully completed EPAs and received a low number of unsuccessful EPAs (i.e. scored a 3 or below on the entrustment scale). The narrative feedback provided by faculty members is largely positive, describing instances where Alex demonstrated exceptional professionalism and strong clinical performance. In the preliminary discussion, these data appear to align with the CCC's perception of this resident as a strong performer.

The evidence presented thus far, consisting largely of assessor-generated rankings of a trainee's ability to perform entrustable activities on a five-point scale, is easily transformed into an aggregate measure that communicates Alex's competency in specific domains. This type of evidence lends itself to the 'follow the numbers' decision-making process valued by this committee:

> The main factors [in decision making] are what we see from the evaluation themselves and not to let our personal knowledge of residents bias us too much, because we have to go with what we are given. We are a committee that is there to see what the summary of evaluation tells us. So, I think we have to really rely on those and be very stringent that we can't deviate from those evaluations by rumor, personal knowledge, all this kind of stuff about a resident. We're just there to look did they make the grade based on the system that we've thrown them into? Yes or no? And then move on. (Committee Member 1)

The presenter then draws the attention of the group to Alex's score on a recent national examination; the score places Alex in the 15th percentile. In previous discussions, standardised examination data had been interpreted by members of this committee as a reliable measure of medical expertise. Given the perceived strength of these data, scores below the 25th percentile necessitated the implementation of specific remediation strategies such as the assignment of an academic advisor or a remedial examination. This process was common across CCCs; examination scores were frequently used to identify outliers and initiate remediation:

> If you're not within the confidence interval or you're within the really low portion of the bell curve, then that is a red flag because everyone else writes that exam in the country [. . .] so there are enough scores that you can say well if everyone is scoring between 35 and 45 and you are getting 16 there's a problem. (Programme Director 1)

In this instance, however, the CCC does not follow this schematic process; Alex's examination score is perceived to be problematic data and thus does not automatically trigger the assignment of an academic advisor. Instead, the CCC members interrogate the data point, contrast it to, and read it against, instances where Alex was observed to demonstrate medical expertise and a strong knowledge base in the clinical workplace. This CCC's deviation from their 'follow the numbers' approach was sparked by an evidentiary disjuncture; the standardised examination score did not align with the CCC's perceptions of Alex or the gestalt of his or her competencies.

As the effortful discussion unfolded, the discrepant data point was further interrogated and the evidentiary value of the examination itself was called into question. The committee member leading the review described at length the variable examination administration practices across North America, and explicated specific factors that may confound or artificially drop a trainee's score. The previously unacknowledged complexity that underlies the examination score and the processes through which it is generated are teased apart by the CCC during a lengthy debate.

The evidentiary value of the score is further problematised when a member draws the CCC's attention to the programme's lack of in-house comparative data:

> One member comments that the post-graduate program has not historically mandated that residents at this stage of training write this exam, and thus, they "have no idea what the expectations are". Given their lack of comparative data, they do not know what scores residents at this training stage ''should be getting''. The scores, overall, for this cohort are similar to those of the cohort above. (Field note 7)

This effortful discussion evolves into a broader debate about the interpretation, weighting and evidentiary reliability of this particular examination for all trainees at this stage of training, given the various unknowns and potential confounding variables. In an interview following this effortful discussion, we asked a member of the committee to describe how standardised examination scores will be interpreted in future meetings:

> We don't particularly know. I myself don't really know what the appropriate comparison is. Who are they being percentile ranked against? We don't actually know that. We're assuming that it's across Canada. But it's a North American test. There's actually flexibility on when you can write it. So, there's a little bit of mystery around that as well [. . .] how can you really compare that? (Committee member 5)

In this instance, a single data point (Alex's examination score) was perceived by the CCC members as problematic and sparked an investigation that ultimately called into question the utility of the examination score itself as an indicator of medical expertise. The discovery of problematic evidence prompted the CCC members to unpack their assumptions about the source of this data point and the socially situated processes through which it is generated. The ensuing discussion was rendered effortful because CCC members were required to wrestle with the disjuncture of the problematic nature of the evidence they were presented with, their interpretive resources that shape how they come to perceive certain types of data as useful evidence, their collective understanding of objectivity and the socially situated practice of decision making. In the story of Alex, the process of effortful discussion prompted the CCC to re-examine the previously assumed objective quality of the examination score and calls into question its evidentiary value in the interpretive process.

Ultimately, the CCC decided to defer its decision about Alex's potential need for academic advising until they acquired additional high-quality data. The decision reached by this committee is illustrative of the two most common responses to problematic evidence identified in our analysis: deferring decision making and asking for more evidence. Further, to ensure the quality and usability of future data collected, this CCC took strategic action and fed recommendations back to the programme to refine the programme of assessment and increase the diversity of tools implemented. Our results suggest that this form of responsive strategic action is a product of CCC engagement with problematic evidence.

Insights

Although problematic evidence often is identified or named as 'problematic' because it does not align with the group's perception of the trainee's performance, it is important to note that problematic evidence, as we have conceptualised it, is not limited to discrepant data – or standardised examination scores that are perceived to be aberrant.

As illustrated in Fig. 1, evidence can be rendered problematic under one or more of the following conditions: when data are absent or missing (i.e. not collected or lost), when performance data are informal and thus lacks credibility (i.e. verbal reports from faculty members, first-hand experiences of committee members), when the tools or practices used for assessment are flawed or ambiguous (i.e. lack of faculty members' consensus around scales or criteria, assessment forms without comment boxes) and when data do not align with the committee's perceptions of a resident or the gestalt of his or her competencies. In the story of Alex, evidence was rendered problematic by

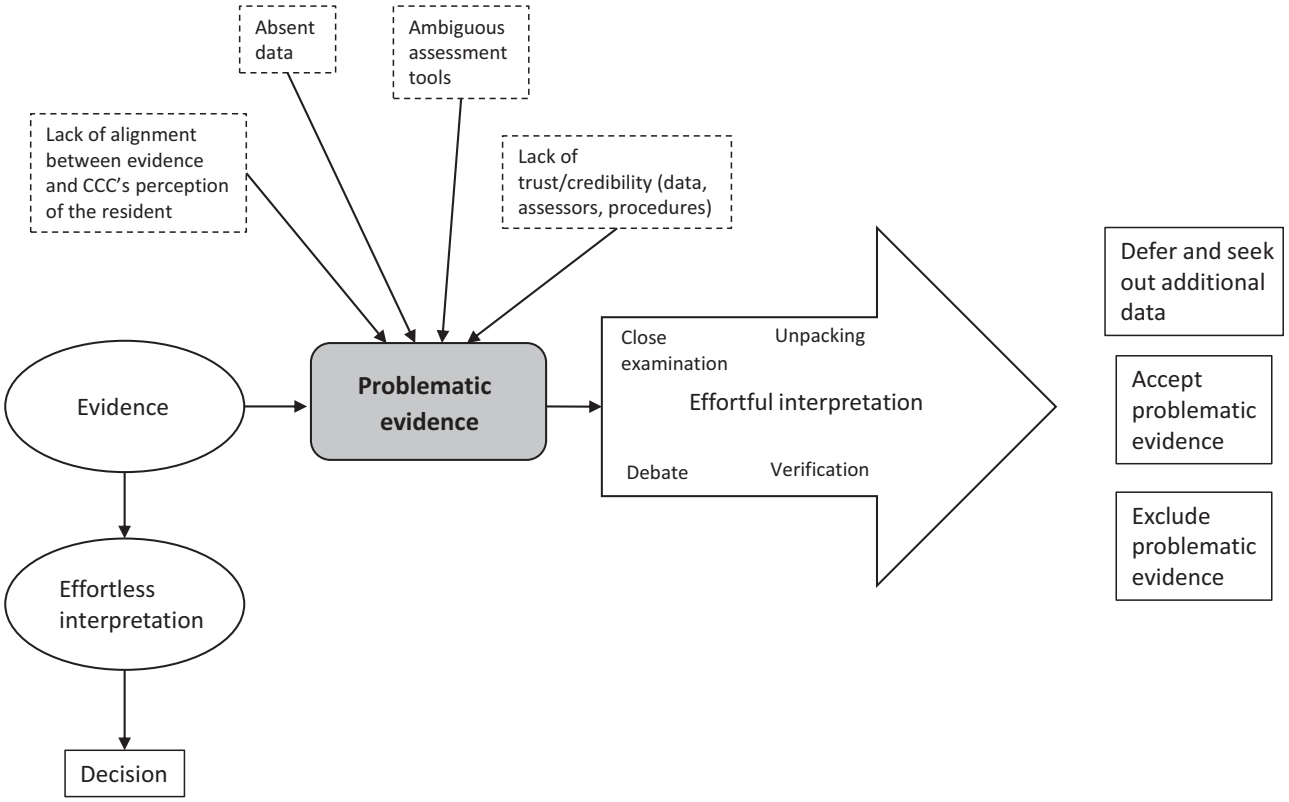

Figure 1 Theoretical model of data interpretation

ambiguous assessment conditions, and the lack of alignment between the group's informal experiential knowledge of Alex and the examination data. As we have illustrated, the identification of this problematic evidence triggered effortful interpretation that involved the close examination, debate and unpacking of the data, which questioned their substantive value.

Our illustrative example revolved around the misalignment of a standardised test score and the committee's perceptions of a trainee's medical expertise; however, many other forms of problematic data existed in our data. For instance, assessment forms that featured both numeric rating scales and narrative comments often were interpreted as being problematic when there was a lack of alignment between the data points. In instances where the narrative and numeric data were in conflict (i.e. positive feedback and a poor rating) the credibility of the assessment as a whole was called into question.

> I think it's not that the information that is provided is not credible, because the people providing it are credible, but I think the pressures to be positive and constructive [in narrative comments], sometimes outweigh the ability to provide truly negative feedback or to really say to someone "you are not where you need to be" [. . .] And the burden of proof certainly lies with the program, to be able to defend the decisions that they are making. That's why I think we all feel it's so important to have the documentation in line with what we are putting on the evaluation. (Committee member 3)

Another iteration of problematic evidence that frequently appeared in our results was variance in how assessors interpreted and applied numeric entrustment or rating scales. This variance made it difficult for CCC members to meaningfully interpret numeric scores and resulted in numerous discussions about whether a '3/5' was *really* a 3.

> There's so much confusion about that entrustment scale, it's crazy. And it was sold as it's so intuitive, it's so easy. There's so much confusion about that - which calls into the quality of what you're analyzing. (Committee member 4)

As the above examples show, CCCs encounter a variety of iterations of problematic evidence, some of which are likely to be familiar to readers with

CCC experience. However, the shape that problematic evidence takes is not the focus of our analytic attention; rather, we are concerned with the tensions revealed by the interpretive processes that problematic evidence necessitates. In the story of Alex, it is the stark contradiction between the CCC's espoused commitment to 'follow the numbers' wherever they may lead and their enacted practice of augmenting the 'numbers' with additional, contextual data that provide an important insight into the processes of CCC data interpretation, namely socially situated data have value in the meaning-making process.

The perceived utility or value of contextual data was articulated across the CCCs in this study, particularly in instances where members struggled to meaningfully interpret a piece of assessment data. Indeed, as a result of their frequent encounters with problematic evidence, our participants appeared to consider, incorporate and even embrace, experiential evidence and first-hand accounts of performance that had previously been rejected as subjective. The rationale one committee chair provided to explain the introduction of 'context' into their interpretive process – data that had previously been excluded for its subjectivity – readily echoes the above insight:

> We don't know if [this new process is] right or not either, but we feel like it's not just an isolated silo of EPAs, and we should be looking at the entire picture, right? And it's hard when you are just looking at stuff on the computer; if you've never worked with that person, it's hard to put a context into it, right? So, we felt that if people can bring context – and that's the whole point of having the mentor person there and getting feedback from them [. . .] We decided as a Competence Committee that we should be overseeing their entire progression as a resident [. . .] it didn't make sense to only look at what was on the computer. (Committee Chair 1)

The value of bringing 'context' into the data interpretation process, to complement numeric performance data and formally submitted narrative feedback, particularly in light of problematic evidence, was echoed by CCC members across the postgraduate programmes studied.

Further, the intensive review of assessment data conducted by the small groups and their numerous encounters with poor quality evidence illuminated the currents gaps within their assessment programmes. One strategy CCCs used to navigate these gaps was to elicit, discuss and consider informal, experiential knowledge from members of the committee:

> I think that there is a lot of data that is lost in the tools that we are using to collect. And some of those experiences that people are adding are their day-to-day experiences, their perspectives and then also, interactions with the residents. But, I think it adds a lot [. . .] it fills the void sometimes in what the [assessment tool] has [. . .] having multiple individuals give their conception of how someone is performing in comparison to their peers, or the expectations that we would have for a certain person at that level of their training, and the ability to kind of talk about that out-loud, and even debate maybe people's performance, I think adds a richness to the evaluation that's better than the [assessment] forms. (Programme Director 2)

Contextual, experiential and informal data shared by members of this CCC and others within the study came to be understood as important sources of evidence that could complement, fill the gaps in and enrich the formal assessment data available for review. This suggests that the professional judgement of the assessors on the CCC was interpreted by members of the committee as an important source of data that could aid in the interpretation of other types or instances of problematic evidence. The progression decisions rendered through this process were both informed by the assessment metrics and tailored to the individual trainee.

DISCUSSION

Underpinning much of the literature on CCCs is a presumption that CCCs can and will engage in a process of data interpretation and synthesis that is both uniform and effortless; a process akin to pixels seamlessly fitting together to form a picture. This ideal model of data interpretation, however, reflects the assumption that the assessment data collected about learner performance will be perceived to be both objective and valid, enabling it to be easily synthesised into an intelligible composite. Our main finding – that CCCs grapple with problematic evidence – calls into question this assumption and illustrates that data (even seemly objective data like standardised test scores) are subject to effortful processes of data interpretation. Ultimately, the findings of this study illustrate that data synthesis is

often not effortless or automatic; data require human assembly, interpretation and judgement.

The question of data interpretation is a worthwhile one, aside from whether the data in question are perceived to be of high quality. This is illustrated in our central example, where we highlighted the discordance between standardised examination data and the committee's understanding of a resident's medical expertise – an example that may be familiar to many readers. This example was purposefully chosen for its resonance and ability to demonstrate the 'common sense' nature of the social phenomenon we've described. Our aim was to make the familiar *unfamiliar* by illustrating and challenging an assumption in the assessment literature: that data supported by validity evidence such as standardised examination data do not require interpretation. As the effortful discussion we describe shows, this assumption does not always hold in practice, particularly when the evidence presented does not align with the committee's perceptions of the trainee's performance.

The results of our study support the findings of other studies that suggest that the socially situated[27] professional judgements of CCC members are vital to both the integrity of the evidentiary review process and the determination of decisions on a trainee's progression.[9,12] In light of this, the effortless data interpretation processes we have described may not be as uncomplicated as they appear to be on the surface, and may operate to obscure limitations of the assessment data on which the decision is based. Although the processes of effortless data interpretation were not the central focus of our analysis, further examination of this socially situated process is well warranted to understand its effects and implications.

In addition to drawing attention to the previously unacknowledged complexity of CCC data interpretation processes, our study contributes to a growing body of literature that suggests that subjective, socially situated data have both a place and value in postgraduate assessment programmes in the era of CBME.[27–29] In particular, we suggest that the phenomenon of problematic evidence is a product of tensions that arise from the interplay between the interpretive frameworks that shape how clinical faculty members come to understand, interact with and value evidence and the nature of data that is produced through human-driven assessment practices. More specifically, the systematic 'follow the numbers' approach that we have described is underpinned by an ethos of evidence-based medicine (EBM) and has a ring of obviousness to it; few physicians would claim that they do not rely on the best available evidence to make decisions. Although on the surface this commitment to 'evidence' over 'intuitive' methods of decision making appears to be both responsible and rigorous, the positivist underpinnings of EBM do not lend themselves to the messiness and socially situated nature of trainee performance evaluation.[19,29,30] In the case of CCCs, the privileging of psychometric and quantitative evidence does not necessarily increase the *objectivity* of the decision, but rather, it *obscures* the inescapable subjective elements that enter into and shape all assessment practices.[29] Given the pervasiveness of EBM as an interpretative framework, the idea that all assessment data have subjective elements is deeply disruptive to assessors' sense of ontological security, a factor that likely contributed to the tensions we observed.

In addition to prompting us to reconsider what 'counts' as evidence, the results of our study have some practical implications. Given their intimate knowledge of assessment data, it has been speculated that CCCs are uniquely positioned to identify gaps or limitations in programmes of assessment.[10,12,16] The results of this study highlight how this occurs in practice. As a result of grappling with problematic evidence, the CCCs involved in the study were made acutely aware of the limitations of their current assessment tools and practices. Further, during meetings, CCCs provided specific and actionable feedback about these identified limitations to their postgraduate programme directors and effected change, including the implementation of faculty training, the introduction of new tools and the refinement of current tools to capture feedback about trainees' performance in non-medical expert roles. As the transition to CBME progresses, it will be important for postgraduate programmes to craft a role for CCCs that harnesses this unique potential for revealing vulnerabilities in assessment processes and programmes. Further research is needed to better understand how the role of CCCs might be optimised to support curricular and assessment design.

Limitations

As social constructivists rooted in post-modernism, our aim is not generalisability: as our data show, reality and truth are multiple and overlapping.[31] Instead, in alignment with the aims of constructivist grounded theory, we aim to provide a robust explanation for the patterns and variations present

in our data rooted within its specific context.[26] We invite our readers to explore the transferability of our theorising to their own unique CCC settings.

CONCLUSIONS

Small groups involved in the review of trainee assessment data should be prepared to encounter evidence that is uncertain, absent, incomplete or otherwise difficult to make sense of, and should openly discuss strategies for addressing these challenges. Such problematic evidence prompts effortful processes of data interpretation, which cannot be eased simply by generating more data with strong psychometric properties. This challenge requires us to grapple with the discrepancies between our dominant interpretive frameworks and the inescapably subjective nature of assessment data and professional judgement.

Contributors: all authors contributed significantly to the design, implementation and analysis of this study. RP participated in data collection and analysis, and led the writing of this manuscript. LL participated in the analysis of data and the development of the coding structure, and provided significant feedback on the drafts of this manuscript. CW participated in the analysis of data, the refinement of the coding structure and the revisions of this manuscript. SC designed and conceptualised this study and participated in data collection and analysis. SMC mentored RP throughout the course of the study and was heavily involved in the analysis of data and the development of the manuscript. All authors support and consent to the submission of this work and agree to be accountable for all aspects of the work.

Acknowledgements: we thank our participants for their participation in this study.

Funding: this study was funded by a 2016 strategic grant from the Royal College of Physicians and Surgeons of Canada.

Conflicts of interest: we have no competing interests or conflicts to declare.

Ethical approval: this study received ethical approval from the Western University Health Sciences Research Ethics Review Board in August 2017.

REFERENCES

1 Holmboe ES, Sherbino J, Long DM, Swing SR, Frank JR. The role of assessment in competency-based medical education. *Med Teach* 2010;**32** (8):676–82.
2 Englander R, Frank JR, Carraccio C, Sherbino J, Ross S, Snell L. Toward a shared language for competency-based medical education. *Med Teach* 2017;**39** (6):582–7.
3 van der Vleuten CPM, Schuwirth LWT, Driessen EW, Dijkstra J, Tigelaar D, Baartman LKJ, van Tartwijk J. A model for programmatic assessment fit for purpose. *Med Teach* 2012;**34** (3):205–14.
4 Bok HG, Teunissen PW, Favier RP, Rietbroek NJ, Theyse LF, Brommer H, Haarhuis JC, van Beukelen P, van der Vleuten CP, Jaarsma DA. Programmatic assessment of competency-based workplace learning: when theory meets practice. *BMC Med Educ* 2013;**13** (1):1–6.
5 Hauer KE, Chesluk B, Iobst W, Holmboe E, Baron RB, Boscardin CK, Cate OT, O'Sullivan PS. Reviewing residents' competence. *Acad Med* 2015;**90** (8):1084–92.
6 Promes SB, Wagner MJ. Starting a Clinical Competency Committee. *J Grad Med Educ* 2014;**6**:163–4.
7 Andolsek K, Padmore J, Hauer K, Holmboe E. Clinical Competency Committees. A guidebook for programs. 2015. https://www.acgme.org/Portals/0/ACGMEClinicalCompetencyCommitteeGuidebook.pdf [Accessed 15 February 2019.]
8 Dickey CC, Thomas C, Feroze U, Nakshabandi F, Cannon B. Cognitive demands and bias: challenges facing Clinical Competency Committees. *J Grad Med Educ* 2017;**9** (2):162–4.
9 Oudkerk Pool A, Govaerts MB, Jaarsma DDC, Driessen E. From aggregation to interpretation: how assessors judge complex data in a competency-based portfolio. *Adv Health Sci Educ Theory Pract* 2017;**23** (2):275–87.
10 Ekpenyong A, Baker E, Harris I, Tekian A, Abrams R, Reddy S, Park YS. How do Clinical Competency Committees use different sources of data to assess residents' performance on the internal medicine milestones? A mixed methods pilot study. *Med Teach* 2017;**39** (10):1074–83.
11 Colbert CY, French JC, Herring ME, Dannefer EF. Fairness: the hidden challenge for competency-based postgraduate medical education programs. *Perspect Med Educ* 2017;**6** (5):347–55.
12 Colbert CY, Dannefer EF, French JC. Clinical Competency Committees and assessment: changing the conversation in graduate medical education. *J Grad Med Educ* 2015;**7** (2):162–5.
13 Friedman KA, Raimo J, Spielmann K, Chaudhry S. Resident dashboards: helping your Clinical Competency Committee visualize trainees' key performance indicators. *Med Educ Online* 2016;**21** (1):29838.
14 Schumacher DJ, Sectish TC, Vinci RJ. Optimizing Clinical Competency Committee work through taking advantage of overlap across milestones. *Acad Pediatr* 2014;**14**:436–8.
15 Dickey C, Cannon B, Thomas C. *"That Resident is Terrific, give her a 3!" and Other Forms of Bias in Clinical Competence Committee Meetings*. Austin, TX: American Association of Directors of Psychiatric Residency Training 2016.

16 French JC, Dannefer EF, Colbert CY. A systematic approach toward building a fully operational Clinical Competency Committee. *J Surg Educ* 2014;**71** (6):e22–7.
17 Kinnear B, Warm EJ, Hauer KE. Twelve tips to maximize the value of a Clinical Competency Committee in postgraduate medical education. *Med Teach* 2018;**26**:1–6.
18 Boscardin C, Fergus KB, Hellevig B, Hauer KE. Twelve tips to promote successful development of a learner performance dashboard within a medical education program. *Med Teach* 2018;**40** (8):855–61.
19 Van Der Vleuten CP, Norman GR, De Graaff E. Pitfalls in the pursuit of objectivity: issues of reliability. *Med Educ* 1991;**23** (2):110–8.
20 Royal College of Physicians and Surgeons of Canada. Competence Committee guidelines - Terms of Reference. 2018. http://www.royalcollege.ca/rcsite/cbd/assessment/committees/managing-competence-committees-e [Accessed 29 January 2019.]
21 Glaser B, Strauss A. *The Discovery of Grounded Theory: Strategies for Qualitative Research*. Chicago: Aldine 1967.
22 King N, Horrocks C. *Interviews in Qualitative Research*. Los Angeles, CA: SAGE Publications 2010.
23 Dey I. *Grounding Grounded Theory: Guidelines for Qualitative Inquiry*. San Diego, CA: Academic Press 1999.
24 Charmaz K. *Constructing Grounded Theory: A Practical Guide through Qualitative Analysis*. London: Sage 2006.
25 Watling CJ, Lingard L. Grounded theory in medical education research: AMEE Guide No. 70. *Med Teach* 2012;**34** (10):850–61.
26 Charmaz K. *Constructing Grounded Theory*. London: Sage 2014.
27 Hodges B. Assessment in the post-psychometric era: learning to love the subjective and collective. *Med Teach* 2013;**35** (7):564–8.
28 Eva K, Hodges B. Scylla or Charybdis? Can we navigate between objectification and judgement in assessment? *Med Educ* 2012;**46** (9):914–9.
29 Ginsburg S, McIlroy J, Oulanova O, Eva K, Regehr G. Toward authentic clinical evaluation: pitfalls in the pursuit of competency. *Acad Med* 2010;**85** (5):780–6.
30 Hodges B, Albert M, Arweiler D *et al*. The future of medical education: a Canadian environmental scan. *Med Educ* 2011;**45** (1):95–106.
31 Lyotard J. *The Postmodern Condition: A Report on Knowledge*. Minneapolis: University of Minnesota Press 1984.

Received 30 October 2018; editorial comments to authors 17 January 2019; accepted for publication 22 February 2019