Translating Outcome Frameworks to Assessment Programs: Implications for Validity
Translating Outcome Frameworks to Assessment Programs: Implications for Validity

Tavares, W.,¹ Rowland, P.,¹ Dagnone, D.,² McEwen, L.,² Billett, S.,³ Sibbald, M.⁴

¹The Wilson Centre and Post-MD Education. University Health Network and University of Toronto, Toronto, On. Canada.
²Faculty of Medicine, Queens University, Kingston On. Canada
³School of Education and Professional Studies, Griffith University, Queensland, Australia
⁴Centre for Simulation Based Learning and Department of Medicine, McMaster University, Hamilton, On. Canada

Abstract:

**Introduction:** Competency based medical education (CBME) requires that educators structure assessment of clinical competence using outcome frameworks. While these frameworks may serve some outcomes well (e.g., represent eventual practice), translating these into workplace-based assessment plans may undermine validity and therefore trustworthiness of assessment decisions due to a number of competing factors that may not always be visible or their impact known. Explored here is the translation process from outcome framework to formative and summative assessment plans in postgraduate medical education (PGME) in Canadian universities.

**Methods:** We conducted a qualitative study involving in-depth semi-structured interviews with leaders of PGME programs involved in assessment and/or CBME implementation, with a focus on their assessment-based translational activities and evaluation strategies. Interviews were informed by Callon’s theory of translation. Our analytical strategy involved directed content analysis, allowing us to be guided by Kane’s validity framework but still participating in open coding and analytical memo-taking. We then engaged in axial coding to systematically explore themes across the dataset, various situations and our conceptual framework.

**Results:** Twenty-four interviews were conducted involving 15 specialties across three universities. Our results suggest (a) using outcomes for assessment frameworks serves as necessary for good assessment but also as an incomplete construct; (b) there are a number of social and practical negotiations with competing factors that displace validity as a core influencer in assessment planning, including implementation, accreditation and technology; (c) validity exists as threatened, uncertain and assumed due to a number of unchecked assumptions and reliance on surrogates.

**Conclusions:** Translational processes in CBME involve negotiating with numerous influencing actors and institutions that, from an assessment perspective, provide challenges for assessment scientists, institutions and educators to contend with. These processes are challenging validity as a core element of assessment designs. Educators must reconcile these influences when preparing for or structuring validity arguments.
Abstract:

Introduction: Competency based medical education (CBME) requires that educators structure assessment of clinical competence using outcome frameworks. While these frameworks may serve some outcomes well (e.g., represent eventual practice), translating these into workplace-based assessment plans may undermine validity and, therefore, trustworthiness of assessment decisions due to a number of competing factors that may not always be visible or their impact knowable. Explored here is the translation process from outcome framework to formative and summative assessment plans in postgraduate medical education (PGME) in three Canadian universities.

Methods: We conducted a qualitative study involving in-depth semi-structured interviews with leaders of PGME programs involved in assessment and/or CBME implementation, with a focus on their assessment-based translational activities and evaluation strategies. Interviews were informed by Callon’s theory of translation. Our analytical strategy involved directed content analysis, allowing us to be guided by Kane’s validity framework, whilst still participating in open coding and analytical memo-taking. We then engaged in axial coding to systematically explore themes across the dataset, various situations and our conceptual framework.

Results: Twenty-four interviews were conducted involving 15 specialties across three universities. Our results suggest (a) using outcomes for assessment frameworks serves as necessary for good assessment but also as an incomplete construct; (b) there are a number of social and practical negotiations with competing factors that displace validity as a core influencer in assessment planning, including implementation, accreditation and technology; (c) validity exists as threatened, uncertain and assumed due to a number of unchecked assumptions and reliance on surrogates.

Conclusions: Translational processes in CBME involve negotiating with numerous influencing actors and institutions that, from an assessment perspective, provide challenges for assessment scientists, institutions and educators to contend with. These processes are challenging validity as a core element of assessment designs. Educators must reconcile these influences when preparing for or structuring validity arguments.

Introduction
A key process within competency based medical education (CBME) is how educators structure assessment of clinical competence. Outcomes/competencies serve to describe features of practice that clinicians must possess/demonstrate to be promoted to independent practice. For example, “entrustable professional activities” (EPAs) are descriptive work tasks that include person-level competencies intended to capture features of practice essential in developing trainees and ensuring their readiness for independence. Educators are tasked with developing, selecting and planning the best combination of assessment strategies to facilitate representation of the complex constructs or achievement of specific outcomes, and/or to inform summative decisions regarding trainee readiness. However, the challenge of this complex task becomes apparent as individuals and organizations translate these lists of outcomes into assessment plans. As outcome frameworks are “translated” – referred to in this case as the active process of generating assessment plans from outcome frameworks – a number of unarticulated influences, assumptions and problematic translations may create misalignments with a foundational tenet of assessment: assessment validity.

The design of assessment plans has typically focused on core features, without having to accommodate large outcome frameworks and/or taking into consideration the variable impact of influences that may be shaping decisions about those plans. For example, item scale development often begins with identifying items or dimensions, stimuli and/or activities that collectively support representation of the construct to be assessed. Other assessment design features have focused on the functional need to overcome context specificity, to identify and then adjust for rater tendencies, to consider the value of assembling diverse points of data, the value of narratives or portfolios, and/or learner roles in assessment. However, the upstream process of translating outcomes frameworks to assessment plans, as well as the type, role and impact of influences, who generates those influences and why, have generally not been explicitly featured in existing frameworks. Assessment validity (i.e., trustworthiness and defensibility of the data) may therefore be undermined without clear guidance on how to identify and navigate these influences.

Using sociological and sociocultural accounts of assessment designs in a CBME context can inform the translation from outcome frameworks to assessment plans. Similar theoretical approaches have been used to explore activities in health professions education, in higher education, in repurposing standards and standardization as learning tools, and in practices of assessment of professional learning. Callon’s theory of translation is one way to understand and explain processes involved in translating from one set of practices (e.g., creating outcome frameworks) to another (e.g., creating assessment plans). Callon’s theory of translation suggests that the act of translation is not just a cognitive exercise, but also social, political and material acts. To translate an idea existing in one context or form (e.g., outcome frameworks) to another (e.g., assessment plans and their implementation and use), individuals and/or groups must convince others of the fit between their idea and issues related to that other context. They must also convince others that the proposed action is a desired and appropriate solution. Throughout this process, individuals must also enroll others in material acts, strategize to keep involved actors (i.e., people, groups/committees
but also material objects such as learning platforms and technologies) engaged in the desired activity and reduce competing demands, and manage any inevitable controversies (REF). These upstream processes may be invisible during the process, not always prioritizing validity, and immutable by the time assessment plans have been created.

Viewed in this way, CBME invites new challenge for assessment science when moving from outcome frameworks to assessment plans. Unlike assessment designs that exist where control over content and context can be exercised (e.g., written tests, OSCEs), the context of CBME and workplace-based assessments (WBA) may require accommodating a range of competing situational factors, a concept that is not new in education and curricular design. For example, Billet demonstrated that workplace curriculum is subject to the affordances and interests of the workplace and actors and their responsibilities within it. Applied to assessment and on the focus of upstream translational activities, suggests features external to assessment design are relevant and extend beyond anticipated events in once assessment plans are enacted in the workplace. Historically, assessment science has de-emphasized these features in place of other threats to validity (REF) (as described above). Given their emphases, sociological and sociocultural theories (e.g., Callon’s theory of translation) help to bring these features into view. If we can understand and illustrate the translational process well, there is a potential to advance effective outcome-based education and valid assessment programs. For instance, the data on which competence committees structure high stakes assessment, or learners use to structure developmental strategies is dependent on the trustworthiness and defensibility of the data generated. (REF) Doing this well, particularly in workplace-based contexts, challenges the assessment community and stimulates diverse approaches (e.g., focusing on raters, tool and faculty development) that often neglect translational activities or evaluation of their impact on validity. Therefore, in this study we asked what shapes the translation of outcome frameworks to formative and summative assessment plans in postgraduate medical education (PGME) in Canadian universities, specifically how and what tools and processes were built and why; the data from which would be used by competence committees. Also, how are these processes justified or evaluated? We positioned this work with a view that translations are cognitive, practical, social and political activities, but that validity should also be considered, particularly when contributing to the culmination of summative decisions about individuals (e.g., for use by competence committees). We aimed to identify and illuminate optimizing strategies and ways of mitigating tensions and challenges that may threaten validity. See Figure 1 for an illustration of a simplified process of assessment design when using outcome frameworks including where Callon’s theory of translation and Kane’s validity and validation framework are relevant.

Figure 1: Illustration of a simplified version of the process of assessment when using outcome frameworks.
Method

Study Overview

We conducted a qualitative study comprising in-depth semi-structured interviews with individual representatives of PGME programs (e.g., program directors), assessment committees, and/or CBME implementation committees, focusing on their assessment based translational activities as well as actual or planned evaluation strategies. Interviews were recorded and transcribed for analysis which involved methods guided by directed content analysis.9 We used both Callon’s theory of translation and Kane’s validity framework to guide our work. Callon’s theory of translation (described above) was used to direct our attention to specific dynamics of the translation process and to shape the structure of our interview guide. This theorization of how translational processes are enacted aligned well with our observations of the undertaking of CBME, in particular the various people, groups and institutions (i.e., academic, professional, clinical, regulatory, frontline) involved. Kane’s validity framework emphasizes validity as a matter of evidence in support of inferences based on data generates in an assessment process. Validation involves the process of identifying and testing assumptions in the assessment process. We used this flexible argument-based framework to direct our attention in the analysis, drawing out possible implications of the translation process for assessment validity. Ethical approval was provided by the University of Toronto in May 2018 (Protocol # 8333). Additional ethics approvals were obtained from the sites involved in this study.

Setting, Sampling and Informants

This study took place at three Canadian university PGME programs whose mandate was to implement CBME under the implementation and accreditation guidelines of the Royal College of Physicians and Surgeons of Canada (RCPS). Selecting three institutions facilitated our intention for both depth and breadth as the three institutional settings permitted some variability in approaches and opportunities to be more broadly informative. All three institutions were involved in the implementation of CBME but their programs have distinct structure and processes, were at different development, implementation and evaluation stages, and in dissimilar cultural/educational contexts. All had CBME implementation leads and had access to education, assessment and program evaluation...
Translating Outcome Frameworks to Assessment Programs

experts. We used purposeful and maximum variation sampling strategies to recruit program directors and/or their delegates and/or committee members, from various specialties who identified as having participated in the development of assessment plans using outcome frameworks as a starting point. Our recruitment strategy involved contacting potential informants by email using internal lists made available to the research team by PGME leadership. We began at the program director level then expanded to include assessment committee members.

Interview Guide

Informed by our conceptual framing, we organized our interview guide into three main target areas: (1) descriptions and contextualization of assessment program development (e.g., how EPAs are being used, how decisions are being made about how EPAs will be used, the use of other non-EPA assessments, how EPAs and non-EPAs are put into practice); (2) exploration of the translation process including current decision-making practices related to the translation of outcome frameworks into specific assessment plans (e.g. how decisions were made, other work that was involved in launching assessment plans, what shifts/course corrections were made in the process of planning and/or implementation); and (3) accounts of program evaluation activities/plans that would support quality and validity of the assessment plans (e.g. formal or informal evaluation activities related to implementation, plans and/or considerations taken to ensure the rigor of assessment decision making). We used open-ended questions with prompts to explore the various moments in the translation and evaluation processes. We also allowed the interview guide to evolve iteratively, and collected socio-demographic details from informants.

Data Collection

Interviews were conducted by a research assistant trained in qualitative research interviewing and members of the research team (WT, PR, MS). The interviewer was decided based on the informant's institution, to avoid potential or perceived conflicts, and availability of the interviewer and informant. Interviews were conducted in-person or via telephone, audio recorded and transcribed verbatim. These transcripts were checked for accuracy and uploaded into qualitative analysis software (HyperRESEARCH™). We continued in a cyclical and iterative process of recruitment, data collection, analysis and team meetings to allow our sampling and interviews to be modified to better address our research aims and to seek threads of clarification, divergence and/or convergence. This process continued until the authors agreed the analysis fit three criteria of rigor: (1) there was sufficient depth to ensure the theoretical claims being made were grounded in empirical material, (2) the claims being made held conceptual plausibility, and (3) the claims had relevance to a broader academic community.

Analytical Strategy

Our analytical strategy comprised directed content analysis, which allowed us to be guided by our conceptual frameworks as described above. Consistent with CBME, we assumed that our informants were structuring assessments for formative purposes, but that data
generated would also be used for summative decisions, obligating a degree of validity evidence in both generating and interpreting data. Given that our intention was to consider translational issues on validity, we used Kane’s framework to sensitize ourselves to where unchecked assumptions may exist in the development and enactment of assessment plans.

After confirming accuracy, research team members first read the transcripts to garner a general impression of the dataset as a whole. We allowed our conceptual frameworks to inform our reading and coding of the data. Using our conceptual framing and philosophical positions, the second reading involved open coding and analytical memo-taking consistent with directed content analysis. Following the open coding, we engaged in axial coding (i.e. establishing linkages) to systematically explore themes across the dataset, various situations and our conceptual framework. Members of the research team met regularly to discuss interim analyses, revisions to the interview guide and triangulation of analyses and positions.

*Researcher Stance*

All transcripts were reviewed and coded by three research team members, each approaching the data from different, but overlapping, conceptual/philosophical positions. These included views informed by clinical, educational, implementation, sociological and validity perspectives. These diverse underlying lenses allowed us to consider the various features associated with translational activities, but with what contextualized practical implications might be. Further, all researchers adopted a constructivist / interpretive stance despite having different locations in relation to the various activities associated with assessment in a CBME context. MS is a researcher and program director involved in the translational process. WT is an assessment scientist with a role in supporting programs in their translational processes. PR is a researcher with an interest in professional learning in workplaces. We considered ourselves as “dis-interested”, meaning we could study the translational process and its potential implications on assessment validity without advocacy for or against any approach, but simply witnessing the perspective of those doing the translation. We approached this work with the underlying assumption that validity (or justification or assessment activities in general) is an inherent consideration in assessment designs and planning. However, we also recognize that this may not be a universally held principle for all involved, as many focus of reliability as an equal concern.

*Results*

A total of 24 interviews averaging 50 minutes were conducted with 16 females and 8 males from 15 different PGME programs across the three Canadian universities (12, 7 and 5 interviews) allowed us to achieve the levels of rigour and coverage specified in our data collection strategy. All interviewees had developed assessment plans from outcome frameworks that had been developed locally or nationally. The programs varied in the degree of local input and support they had, in establishing assessment programs, from none at all (i.e., program directors working independently) to access to assessment committees, education and assessment scientists. Also all interviews reported having support of the accreditation body, and had established competence committees for which the assessment
data generated by their assessment plans would be used for progress, promotion and summative decisions. Some of the interviewees held more than one role relevant to this study, for example holding a position on competence committees and/or assessment committees, in addition to being a program director.

The findings presented here attend to the process and consequence of moving from outcome frameworks to assessment plans and opportunities for validity. This movement occurs as programs (e.g. individuals, groups, organizations) prepare for or enact assessment plans against an already established or interim outcome framework that intends – in this case – to represent competence of specialist physicians. Through discussions with those responsible for the translation of outcome frameworks, this section provides findings related to how this process creates and takes place in a context that make assessment validity difficult, uncertain or de-emphasized. We now discuss how validity must compete and be negotiated with other CBME-related priorities in resource-limited environments.

*Positioning outcomes frameworks as necessary for good assessment but also as an incomplete construct*

Individuals, committees/groups and organizations have been successful in establishing a fit between outcome frameworks and assessment validity, to the extent that their connections were not challenged by the informants in this study. That is, the construct of clinical competence for a given specialty could be defined - a requirement for validity - by the collection of competencies/EPAs. Also, that accumulating evidence of the achievement of these artefacts solves the “problem” of having to make defensible decisions about progress, achievement or readiness. The use of outcomes served as necessary for good assessment in CBME. This view of the suitability of outcomes frameworks as assessment items/domains, was informed by authoritative groups indirectly (or sometimes directly) indicating to actors to trust in template assessment tools, often with little or no translation at all. Illustrative quotations informing this trust are:

*P21: We basically translate, essentially verbatim, the basis of the EPA... would be the EPA verbatim and the milestones verbatim. With the comment section at the bottom, and all the contextual variables, in dropdown menus above.*

*P11: I mean - I'd say it wasn't decided by [accreditation body] because I would say it was the consensus of all of the people in that [specialty] working group, which includes program directors across the country....*

In this way, collections of outcomes as assessment tools, that is moving EPAs to make up items on assessment tools with very little or no translation at all, are held to be inviolate and relied upon. However, actors (e.g., program directors, assessment committees) were enrolling in these assessment plans without direct authority over them. This lack of discretion was both accepted and, at times, a difficult negotiation between groups (e.g., users, accreditors, implementers). As a result, this negotiation was associated with some displaced accountability and uncertainty regarding what the enactment of established tools might mean for validity. Issues of validity were often entrusted to those making the
recommendations, despite flexibility in how assessments can be built and the option of maintaining historical assessments (i.e., those prior to implementation of CBME).

Further, while validity often begins with an intended construct, outcome frameworks were identified as incomplete for this role. Informants described the need to supplement translated outcome frameworks, suggesting some uncertainty in how constructs are to be defined, for example:

_**P13:** We call it the other bucket. You know not everything belongs in an EPA or, you know...not everything is EPA-able. And so there’s still lots of other things that people do that don’t neatly fit into EPAs, and so how do you still capture that information which is really important...[specialty] called it the special sauce...But it’s [EPAs] are not the be-all and end-all. Physicians aren’t only EPAs. And there was lots of stuff that didn’t fold neatly into there._

Informants seemed to be indicating that translated outcome frameworks had limitations either in the outcome framework itself or in its translation to assessment frameworks. As such, many informants reported being reluctant to abandon their previous assessment tools, preferring to integrate EPAs into a pre-existing system of in-training assessments.

**Negotiating through competing influences on assessment plans**

Assessment plans are one of many features in CBME and so moving from outcome framework to assessment plans involved a number of influencers (e.g. actors and situational factors), other than validity, broadly grouped as implementation (i.e., including change management), accreditation and technology. This led to streams of activities with competing demands or tensions. That is, these influencers intentionally and unintentionally strengthen their own or weaken others’ positions and, validity is represented not as the core influencer in the translation process, but as one among many.

Implementation, refers to the need to enact CBME in a way that facilitates achievement of intended principles (e.g., provide trainees opportunities to achieve EPAs, identify learner progress or those in difficulty) and that how this is enacted or prioritized is a choice for those involved. This obligation to implement CBME as an activity to be accomplished, is enacted sometimes not to the exclusion of validity, but as a higher priority. At times implementation involved a recognition of the shift and buy-in of educators that would be required and how this too must be prioritized, as illustrated below.

_**P15:** ...I did some initial work just to get stuff out there and get people doing something. Because I just fundamentally thought, because we had the deadline and we had to get this implemented...I was looking at it from a very pragmatic point of view of how can I get people to do this? How can I get people to buy in? How can I make it so that it's easy for them to do? Which I think is a really important feature because if you roll out a system that is clunky and ugly and doesn't speak to people and they don't feel it's relevant, they're not gonna use it, and they're not gonna do it, or they're gonna do it poorly._
These real-world issues appear to be strong influences on those enacting assessment plans. Implementation decisions can exist in tension with – even if only at times or for a period of time – with validity.

Accreditation, as an influence on decision-making, refers to the responsibility to comply with accreditation requirements which at the time for many were perceived to still be fluid, uncertain and not always or necessarily a validity issue. This focus on accreditation meant balancing, relying on or complying with the guidance provided by the accreditor for assessment plans, as illustrated here.

*P22: ... I don't think anyone in our institution was hoping to go rogue and do um something that unique... not that we have a fundamental disagreement with what the [accreditor] was suggesting, in fact we agree...it's probably, you know, a good strategy, makes sense...but there’s just being knowledgeable around the scholarship [and] progress on best practices and assessment. So I think there's what you have to do because you're being accredited and then there's what you should do because it's the right thing to do based on what you know about best practices in assessment.*

Many of those acting to support accreditation served to influence how actors proceeded with assessment plans, as depicted in the following quotation.

*P22: The latitude for creativity is, is probably a little narrower and the whole kind of curriculum and assessment design, has now just fallen back to the [accreditation body] and the relevant specialty committees. And we’re just sort of pointing programs back to what's coming out of the [accreditation body] as documents and those are the tools that you're going to be using.*

There is a strong and important institutional imperative to comply with externally-derived statements for regulatory needs, yet an accompanying concern this may restrict some discretion or need to attend to validity implications.

Technology, closely linked to both implementation and accreditation, refers to the widespread intention to have assessment processes moved on-line. In this instance, assessment tools and reporting were limited to technological capabilities and this often was a more powerful determinant in how assessments were enacted than considerations for validity, as illustrated below.

*P18: We had a system that would allow programs to design an EPA assessment tool that didn’t use the milestones as the questions. But then in the background, we would program the system to be able to map the questions that were used to the relevant milestones. And that was okay for a couple of programs, but once we started to launch more local, it became very obvious that the programming and customization required to do that, was going to be, um, unsustainable, and in fact, overwhelming. So, um, so we had to change the system we were using, not the IT system.*
In this case a physical constraint – a particular technology – limits the range of options and potential for responses and discretion by those enacting assessment plans. Such constraints, therefore, have a strong influence on what assessment approaches and practices can be enacted within constraints of an institutional setting. This extends to what can constitute validity.

*Positioning validity as threatened, uncertain and/or assumed*

We examined decision-making processes using Kane’ validity framework to organize where unchecked assumptions may be in the translation into and enactment of assessment plans, as well as where gaps may be. Validity threats emerge in both the translation process (e.g., developing and using standardized tools, multi-purposing outcomes) and the enactment of assessment plans (e.g., access to events, rater and learner behaviors, workplace affordances).

In addition to issues of construct representation (described above), validity frameworks provide an opportunity to explore unchecked assumptions that serve as opportunities to support the generation of validity claims. Many assumptions can be addressed with data that inherently exists as a result of the implementation. However, as illustrated below, many informants reported being unprepared/ill-equipped to examine validity evidence, and that opportunities have not been available or attention to generating evidence of validity has not yet been prioritized.

P15: *For many years, it was just get it done, get something out there that you can have going by [launch date]. Could it be better? Of course. I think it's really hard to know, and that's actually what we are struggling with now. How do you measure these things? I don't know. This is a challenge and a struggle...I'm not sure how to answer that (referring to a question on how assessment plans are evaluated). Hopefully you can do that for me. I would really like to know how to evaluate it, 'cause I think it's working, but I find it's really hard to, um, to study and prove its working.*

P18: *Um, what we haven't done is really, uh, attempted to design or recognize anything that would sort of get at how valid any of this is...you know, always how difficult, that aspect of it's gonna be, and we're no further ahead than anybody else, in that way.*

This challenge of attempting to evaluate assessment plans seemed to be complicated by how few organizations had either planned or implemented an evaluation of their implementation efforts. Many deferred this to postgraduate departments, as an institutional process after implementation, rather than an integral part of the assessment system.

Displays of validity are distributed across “institutions”, leaving validity arguments poorly developed if at all. Accountabilities for validity are either displaced, reduced or put aside while other priorities are attended to. Accountabilities are often delegated in more than one...
direction. A number of surrogates are reported being used to infer some degree of defensibility/validity, often rooted in best practices:

P15: I can say our competence committee has met multiple times, and we've looked at the data; we've looked at the trends. We're- they're doing okay. Like they're getting the right number of assessments done. We're making sure that they're getting a breadth of assessors in different clinical situations. So I feel like we're going through the process of CBD [Competency By Design, the accreditation body’s version of competency based education] the way it's intended to be going. And that in many ways is a marker of success. Although from a scholarly point of view you could say we're- well, maybe we're just sweeping it under the rug and saying that they're doing okay and advancing them...I don't know.

P20: I think they are finding things out earlier. We're hearing that, that problems are emerging earlier. Um, and I think with more assessments, um, I think it’s just brought that conversation up to the forefront.

In summary, concerns about validity were not the primary influencer when translating outcomes. At best, it remained assumed, based on surrogates or that it could be transferred without further testing of assumptions or validation as evidenced by the systematic absence of evaluations. To be fair, for many, extensive evaluations of validity was not yet a possibility due to the stage of implementation. However, validity was de-emphasized or displaced in these contexts as well. Construct representation meant reaching beyond EPAs or their competencies, and threats appeared in plans following translations, many of which were assumed to be mitigated through competence committees or use of programmatic assessments. As result of many of these issues, validity evidence associated with the data generated for competence committees, may be assumed, unchecked and with uncertain accountabilities.

See table 1 for a summary of themes with key points, and theoretical and practical implications.

INSERT TABLE 1 HERE

Discussion

CBME serves to inform a set of educational aims supporting curriculum strategies and also assessment plans intended to support decisions regarding trainee readiness for unsupervised practice. The structuring of assessment plans in this context often includes the translation of outcome frameworks against competing factors that may unknowingly threaten validity claims. Left implicit or unattended to, the data on which decisions are ultimately made about trainees may be undermined with significant downstream implications for trainees, institutions, patients, and the public. In this study, we explored the translation process from outcome framework to formative and summative assessment plans by interviewing CBME leaders within three universities in Canada. Our results suggest (a) using outcomes for assessment frameworks serves as necessary for good assessment but also as an incomplete
construct; (b) there are a number of social and practical negotiations with competing factors that displace validity as a core influencer in assessment planning, including implementation, accreditation and technology; (c) validity existing as threatened, uncertain and assumed due to a number of unchecked assumptions and reliance on surrogates. See Figure 1 for an illustration of these findings.

**Figure 1:** Summary of process and findings related to the translation of outcome frameworks to assessment plans.

Assessment planning in a CBME context begins with outcome frameworks that pass through a translational process which varies in strategies and degree across the three universities included in this investigation. Others have described how outcome frameworks may not translate neatly into assessment frameworks.\(^4\,16\,17\) Some explanations have focused on the quality of the tools or unprepared faculty;\(^18\) or that it is conceptually flawed to start with outcome frameworks because they are socially constructed or poorly aligned with more grounded conceptualizations of competence.\(^4\,17\) When moving from outcome framework to assessment plans, Callon’s theory of translations suggests a number of additional mechanisms may be relevant.\(^10\) The findings of this study suggest that the translational process unfolds (as Callon would suggest) in social, practical/material, theoretical and political contexts and the emergent process shapes eventual assessment plans. The role of social conditions has been highlighted before in educational contexts, including CBME.\(^5\) For instance, researchers acknowledge the importance of specifying educational outcomes and having intentional plans to achieve those outcomes, but demonstrate how contexts and participating actors can create something very different.\(^6\) That is, underlying educational principles may be transferred, transformed or abandoned as a result of contextual (i.e., social and practical) features that are in competition with curricular goals.\(^19\,20\) As a result, Moore\(^21\) described curriculum as experienced, emergent and evolving, as part of the real world with competing priorities and not as a static codified and sterile planned process. The results of this study suggest that in an analogous way, translational activities and assessment plans are also susceptible to a number of influences (i.e., views and priorities of actors and institutions) that leave validity more intended than guaranteed or certain.
Further, in our study, we learnt of threats to validity while also identifying (untested) attempts to mitigate these threats. For example, outcome frameworks were identified as incomplete or as an insufficient construct in their current form. This may support Lurie’s earlier arguments regarding how outcome frameworks may not translate neatly into assessment frameworks.4,16,17 Additional processes were necessary and included to support defensible claims regarding competence. “Scoring” threats such as faculty and learner behaviors, or “Generalization” threats such as difficulty accessing clinical events in sufficient quality/quantity are being met with attempts to mitigate them (e.g., faculty development, use of simulation as a supplement) but remain largely assumed in regards to impact on validity. For instance, there are a range of issues associated with the reliability of assessment practices that arise from judgements and decisions being made in clinical practice as these circumstances cannot be controlled in the way that assessment practices in educational institutions can be to try and support reliability. Other inferences including “Extrapolation” or “Implications” evidence had not yet been an actioned focus nor explicitly planned. Plans to include examination of assumptions and to structure validity arguments are de-emphasized or not considered in place of other competing priorities,22 and/or because validity is assumed and/or attributed to surrogates (e.g., number of assessments completed, identification of learners in trouble).

Assessment scientists and educators in medical education have perhaps not had to contend with these kinds of influences that position validity not as the core influencer in assessment planning. Common guidelines may de-emphasize how to consider other social influences that compete with assessment validity in the planning or evaluation process, beyond feasibility, cost or acceptability.23-25 But validity or validation may be as much a social process (we mean this to be distinct from a social accountability26) as it is an empirical and logical one, that is shaped by context, actors and institutions. In this study we found, for example, wanting or needing to align with the interests of other stakeholders (e.g., implementation, accreditation, technology), displacing of accountabilities and other material influences were shaping assessment designs. This was not always, or necessarily, in ways that would provide the best opportunities for validity. Assessment planning becomes additionally about social interactions and negotiations than pure assessment science. An implication then is to ask what then, should we now make of assessment designs when assessment is but one priority among many across stakeholders with standing in a complex intervention like CBME? Every time a new actor is enrolled into the translation process, they can potentially disrupt it, obligating us to figure out how any competing imperatives are to be addressed. Assessment guidelines, therefore, may need to be reframed to account further for these issues not to compromise on but to ensure sufficient validity arguments can be made when these issues arise. In light of these challenges but also to avoid problematic compromises, validity and/or validation may need to be given sufficient space especially when compromises are necessary. Until then validity may remain assumed using the application of “best practices”, the use of surrogates, and/or by displacing accountabilities (i.e., validity is a “they” responsibility).

Limitations
This study should be considered in the context of its limitations. First, our sample represents three Canadian universities and we recognize the situated nature of this context. As with other qualitative studies, our intention with this study is to have conceptual transferability and we do not suggest empirical generalizability. Second, given the relatively early status of CBME in Canada, the system may not have had the advantage of time or testing of validity arguments through appeals or failures, and may therefore still mature. We recognize that this study captures a particular moment in time. Our intention in conducting this study, this early, was to stimulate these validity considerations at an early stage.

**Conclusions and Future Directions**

CBME is complex intervention that begins with an outcome framework which is then translated into many activities broadly grouped as teaching, learning, and assessment. This translational process takes place in a complex social and practical environment that, from an assessment perspective, provides new challenges for assessment scientists, institutions and educators. This environment is challenging validity as a core influencer in assessment designs. Other stakeholder interests (e.g., implementation, accreditation, technology) and contextual features such as the (de)valuing of and access to validation opportunities, are at play which assigns validity a unique and certainly not always prioritized or even feasible assignment in the process. Validity and validation struggle to compete for space or may be de-emphasized, displaced or just assumed, at least at this point. As a result, there are some unchecked threats and intended mitigation strategies (e.g., some faculty development, seeking large amounts of data) are more assumed in terms of their impact on validity than tested, leaving validity arguments difficult to create, implicit and risky. Moving forward, first where assessment is involved and contributes to decisions about readiness or competence with significant downstream effects for the learner/candidate, public/patients and institutions, validity and validation may need to be prioritized or given more standing when implementing assessment features of CBME. Second, future research from multiple perspectives (i.e. perspectives of trainees, perspectives of members of competence committees) is necessary to provide an even more comprehensive understanding of these processes and their potential effects. Finally, the context of CBME is causing researchers to propose ways of accounting for the complexity of context. Therefore, in a similar way, because translations involve negotiations, assessment researchers may need to explore and report on how validity frameworks are to account for the contextual realities in which assessment in medical/health professions are to be implemented and enacted. In explicating these translational processes in the context of CBME planning and implementation, shifts toward intentionally integrating these issues, rather than ignoring or competing against them, may be a way of optimizing strategies and ways of mitigating tensions and challenges that may threaten validity. This may also require new partnerships, new forms of collaboration, and new insights from other forms of social science that have not yet been fully appreciated within the domain of assessment or validity science. Until then, translational processes may remain implicit and validity of assessment programs in CBME risky.
Author Contributions
WT, PR, MS and SB provided the initial conceptualization and design of this study. WT, PR and MS led the acquisition and analysis of data. All authors contributed to the interpretation, interim analyses and refinements. WT drafted the initial manuscript with all authors making significant intellectual contributions to subsequent drafts. All authors approved the final version to be published.

Acknowledgements
The authors of this study would like to thank the PGME leaders for their support as well as the participants for their insightful contributions.

Source of Funding
This study was generously supported by the Royal College of Physicians and Surgeons of Canada.

Conflict of Interest
None.

Ethical Approval
Ethical approval was provided by the University of Toronto in May 2018 (Protocol # 8333). Additional ethics approvals were obtained from the sites involved in this study. Anonymity of the participants was protected by removing any identifiers, including ensuring any statements could not be traced back to a particular institution, program or individual.
References


