Hidden in plain sight

Valorization

1. **(Relevance) What is the social (and/or economic) relevance of your research results (i.e. in addition to the scientific relevance)?**

The results from this research can have potential economic advantages for medical schools and residency training programs. As described in Chapters 2 and 4 we found that it is possible to rank-order internal medicine residents (from highest to lowest) based on an analysis of their supervisors' assessment comments and that this rank-ordering is extremely reliable. Further, we determined in Chapter 4 that this high reliability can be shown even after only 3 months of residents' clinical rotations, with 2 faculty members analyzing and rank-ordering – far exceeding the results based on numeric scores alone.
Residents ranked as lowest in their cohort(s) can therefore be identified much earlier using this method when compared to looking only at their numeric scores. The low ranked residents can then be flagged for closer attention, follow-up or even remediation if necessary. This would not be possible using the numeric scores until much later – perhaps at 6 months or even the end of the year. Thus there is the potential – with relatively modest time and effort – to act early and get residents on track for success. If remediation is needed it can begin immediately and this can save time and effort at the end of the program.

Furthermore, since we also found (in Chapter 6) that senior residents in the same Internal Medicine could analyze and rank-order written comments as reliably as faculty could, a senior resident could be paired with a faculty member for the task described above, which would further decrease the time and costs involved. It would also have the added benefit of engaging senior residents in an educational process, which could allow them to reflect on their own assessments and which also might spark an interest in educational development or scholarship.

Another benefit of being able to rigorously use written comments for assessment is that it helps avoid the reductionist approach of numeric scoring. When using only numeric scoring and checklists the clinician-supervisor must observe a performance and then mentally translate that into one or more of several possible competencies (e.g., did the performance relate to physical exam skills, communication or professionalism?) following which they must choose a score from 1-5 to represent their assessment. More nuanced and subtle information can get lost in these multiple translations, which affects the authenticity and validity of the assessment process. On the other hand, if the performance can be captured in words, using natural language, it is likely that more "raw" information or data will be preserved. This assessment language can be used to create more robust documentation (beyond numeric scores), which can be critically important to support (or defend) a learner's remediation or dismissal from a program. This is a notoriously difficult, time consuming and expensive process, and any developments to reduce unnecessary time and costs will be of great value.

2. (Target groups) To whom, in addition to the academic community, are your research results of interest and why?

One important target audience is the promotions committees at medical schools and residency programs. These committees are becoming more prevalent as medical education shifts towards a competency-based (rather than strictly time-based) approach. Promotion committees assess a learner's progress and trajectory and decide whether or not they are ready to move to the next stage of training. Since much of the data collected is in the form of words (assessment comments) our research would be of great interest in that it can lead to a more systematic and rigorous method for analyzing these comments. Our findings also provide evidence to support the validity of such an approach, which is critical in ensuring buy-in and in case evidence is needed for an appeal.

The approach we developed – and the new knowledge generated regarding the value of written comments – will also be very helpful to organizations such as the Royal College of Physicians and Surgeons of Canada. As mentioned, with the shift to CBME there will be many more assessments gathered on each resident, much of which may be in the form of written comments. The Royal College also still promotes the use of In-Training Evaluation Reports (ITERs) which were the subject of all of the studies comprising this thesis. Our findings therefore have great potential to inform and guide the RC in determining best practices for writing and analyzing written comments.

Finally, the resident learners themselves are an important target audience. From residents' perspectives, we found that assessment comments may be preferable to numeric scores, or at the least they are considered necessary and complementary (see Chapter 6). In multiple studies researchers have found that resident learners often express frustration about their assessments, reporting that they are not useful to their development. Our findings suggest that if the assessments include more and "better" comments that residents will find them more authentic and helpful. This may lead to better engagement in their assessment process and more "buy-in" with regards to their own feedback. Further research would have to be done, but this system of assessment could have effects on residents' self-directed learning and self-assessment abilities, which are very important to ongoing maintenance of competence.

3. (Activities/Products) Into which concrete products, services, processes, activities or commercial activities will your results be translated and shaped?

Our findings, along with the suggestions we recommend in the Discussion chapter, can be immediately implemented and tested for validity and feasibility. In Internal Medicine, at least in Canada, the writing and interpretation of assessment comments about residents seems fairly universal; that is, internal medicine attending physicians from

across the country came to the same judgments about residents based on reading their assessment comments. We believe that at least in Internal Medicine in Canada it will be possible to transfer these findings to programs across the country, who can begin to use assessment comments in a systematic way. It is important to note, however, that the process and model for assessing comments may not be immediately or directly transferable to other programs, such as surgery or psychiatry, or in other settings or countries, and would have to be rigorously tested prior to use. It would have to be shown that comments from these domains can be interpreted reliably and with similar predictive value when compared to our findings in Internal Medicine. Any program, however, can take our published protocol and easily adapt it for their own research.

Another demonstration of the translation of this research is in its dissemination. For example, Chapters 2, 3 and 5 have already been published in the peer-reviewed literature (Chapter 2 is also available free online from the journal) and have received approximately 20 citations so far. Chapter 5 just came out in print yet has been read nearly 100 times so far on researchgate.net, indicating that it is gathering significant interest. I have already spoken publicly about my research at other universities in Canada and have more public speaking engagements scheduled in the next few months, all of which will serve to disseminate and promote our findings more broadly.
Chapters 4 and 6 are currently under editorial review at two different journals.

A further indication of the potential translation of this work can be found by considering the interest this research has garnered at the organizational level. For example, I have been invited to sit on a newly formed research advisory committee at the Medical Council of Canada, the organization that develops and administers the examinations required prior to obtaining a medical license to practice in Canada. This new committee is tasked with advising the MCC, it's psychometric and research committees about ongoing and future assessment processes necessary to enable the MCC to meet its important goals. My participation will provide an important venue for translating our research findings into practice and also in developing and conducting new studies to further refine and test our findings in multiple programs.

4. (Innovation) To what degree can your results be called innovative in respect to the existing range of products, services, processes, activities and commercial activities?

Our research is innovative because it is the first to show, in a systematic and rigorous way, that written assessment comments about medical trainees can be highly reliable and useful for assessment. Our findings bolster the validity argument supporting the use of such assessments for determining the competence of our trainees. Prior to this, nearly all of our assessments in medical education have been based on numeric scores, which have their strengths and important uses but also have many limitations. This

research therefore advances the field in a way that could have broad implications for any other program that uses workplace based assessment.

5. (Schedule & Implementation) How will this/these plan(s) for valorization be shaped? What is the schedule, are there risks involved, what market opportunities are there and what are the costs involved?

Dissemination of the novel research in this thesis has already begun. In addition to the publications noted above, the thesis book will also be published and publicly available in the fall of 2016. I have begun speaking about this research at local and national conferences and events. In terms of uptake, it will of course take time for these new ideas to be adopted by others but we hope that our research will be replicated in other disciplines. When I present this work publicly I always invite interested researchers or stakeholders to contact me if they are interested in our protocol and I have offered assistance in setting up similar studies elsewhere. As it is quite new it is not yet clear how long it may take until replication studies can be completed, but there is great potential in the next one to two years for this to occur.

At this point it is difficult to foresee immediate market opportunities related to this research and thus it is not possible to estimate costs. However, one relative risk is worth noting as we consider shaping valorization plans and that relates to the prevailing quantitative culture at most institutions. As mentioned in this section and throughout this dissertation (see especially the Discussion section) our medical education systems have been based almost exclusively on numeric scoring systems for assessment. The administrative and regulatory systems rely heavily on gathering and reporting numbers and scores, and are not optimally structured to allow for the collection and interpretation of qualitative data. For new systems – based on qualitative assessment – to not only be implemented but to succeed and reach their full potential will require a significant shift in mindset and culture. This can take significant effort and time but we are confident that this effort will ultimately be fruitful.