Canadian Medical Education Journal
Revue canadienne de l'éducation médicale

If we assess, will they learn? Students' perspectives on the complexities of assessment-for-learning

Si nous les évaluons, apprendront-ils ? Le point de vue des étudiants sur la complexité de l'évaluation pour l'apprentissage

Valérie Dory, Maryam Wagner, Richard Cruess, Sylvia Cruess et Meredith Young

Volume 14, numéro 4, 2023

URI : https://id.erudit.org/iderudit/1106727ar
DOI : https://doi.org/10.36834/cmej.73875

Aller au sommaire du numéro

Éditeur(s)

Canadian Medical Education Journal

ISSN

1923-1202 (numérique)

Découvrir la revue

Citer cet article

Dory, V., Wagner, M., Cruess, R., Cruess, S. & Young, M. (2023). If we assess, will they learn? Students' perspectives on the complexities of assessment-for-learning. *Canadian Medical Education Journal / Revue canadienne de l'éducation médicale*, *14*(4), 94–104. https://doi.org/10.36834/cmej.73875

Résumé de l'article

Introduction : L'évaluation peut influencer positivement l'apprentissage mais la conception de dispositifs d'évaluation pour l'apprentissage efficaces s'avère difficile. Nous avons implanté en formation prédoctorale un système obligatoire d'évaluation pour l'apprentissage comprenant une évaluation en milieu clinique des compétences transversales et un test de rendement progressif, puis évalué ses effets.

Méthodes : Nous avons mené des entretiens semi-dirigés avec des étudiants en troisième et quatrième années de médecine à l'Université McGill pour explorer la manière dont le système d'évaluation avait influencé leur apprentissage au cours de la troisième année. Nous avons effectué une analyse thématique, informée par la théorie, des données.

Résultats : Onze étudiants ont participé. Les résultats indiquent que l'évaluation a influencé leur apprentissage par le biais de plusieurs mécanismes différents. Certains d'entre eux nécessitaient une implication faible de la part de l'étudiant, comme l'identification des objectifs à atteindre (feed-up), l'apprentissage amélioré par les tests, la recherche d'informations après un examen. D'autres exigeaient une implication importante (par exemple, étudier pour les tests, sélectionner les évaluateurs pour obtenir une rétroaction de qualité, mettre à profit la rétroaction). L'implication des étudiants était modulée par leur perception des avantages et des inconvénients de s'impliquer, et de la crédibilité du système. Cette dernière était influencée par les objectifs-en-contexte des étudiants: devenir un bon médecin, contribuer à l'équipe soignante, réussir les épreuves d'évaluation.

Discussion : Notre système d'évaluation n'a pas réussi à impliquer suffisamment les étudiants que pour réaliser son potentiel. Nous abordons les défauts inhérents au système ainsi que les facteurs externes qui ont entravé l'implication des apprenants. Pour implanter efficacement un dispositif d'évaluation pour l'apprentissage, les concepteurs d'évaluations devraient optimiser les mécanismes qui sont faciles à contrôler et être prêts à s'investir dans un important travail de collaboration pour changer les cultures d'apprentissage.

Canadian Medical Education Journal

Brief Reports

If we assess, will they learn? Students' perspectives on the complexities of assessment-for-learning

Si nous les évaluons, apprendront-ils? Le point de vue des étudiants sur la complexité de l'évaluation pour l'apprentissage

Valérie Dory,[1,2,3] Maryam Wagner,[4] Richard Cruess,[4] Sylvia Cruess,[4] Meredith Young[4]

[1]Department of General Practice, Faculty of Medicine, Université de Liège, Liège, Belgium, [2]Department of Medicine and Centre for Medical Education, Faculty of Medicine, McGill University, Quebec, Canada; [3]Institute of Health Sciences Education and Academic Centre of General Practice, Université catholique de Louvain, Brussels, Belgium; [4]Institute of Health Sciences Education, Faculty of Medicine and Health Sciences, McGill University, Quebec, Canada

Correspondence to: Valérie Dory, Avenue Hippocrate, 15 (B23) - CHU Sart-Tilman; 4000 Liège; Belgium; phone: 32.4.366.42.78; email: valerie.dory@uliege.be

Published ahead of issue: Apr 6, 2023; published: Sept 8, 2023. CMEJ 2023, 14(4). Available at https://doi.org/10.36834/cmej.73875

Abstract

Introduction: Assessment can positively influence learning, however designing effective assessment-for-learning interventions has proved challenging. We implemented a mandatory assessment-for-learning system comprising a workplace-based assessment of non-medical expert competencies and a progress test in undergraduate medical education and evaluated its impact.

Methods: We conducted semi-structured interviews with year-3 and 4 medical students at McGill University to explore how the assessment system had influenced their learning in year 3. We conducted theory-informed thematic analysis of the data.

Results: Eleven students participated, revealing that the assessment influenced learning through several mechanisms. Some required little student engagement (i.e., feed-up, test-enhanced learning, looking things up after an exam). Others required substantial engagement (e.g., studying for tests, selecting raters for quality feedback, using feedback). Student engagement was moderated by the perceived credibility of the system and of the costs and benefits of engagement. Credibility was shaped by students' goals-in-context: becoming a good doctor, contributing to the healthcare team, succeeding in assessments.

Discussion: Our assessment system failed to engage students enough to leverage its full potential. We discuss the inherent flaws and external factors that hindered student engagement. Assessment designers should leverage easy-to-control mechanisms to support assessment-for-learning and anticipate significant collaborative work to modify learning cultures.

Résumé

Introduction : L'évaluation peut influencer positivement l'apprentissage mais la conception de dispositifs d'évaluation pour l'apprentissage efficaces s'avère difficile. Nous avons implanté en formation prédoctorale un système obligatoire d'évaluation pour l'apprentissage comprenant une évaluation en milieu clinique des compétences transversales et un test de rendement progressif, puis évalué ses effets.

Méthodes : Nous avons mené des entretiens semi-dirigés avec des étudiants en troisième et quatrième années de médecine à l'Université McGill pour explorer la manière dont le système d'évaluation avait influencé leur apprentissage au cours de la troisième année. Nous avons effectué une analyse thématique, informée par la théorie, des données.

Résultats : Onze étudiants ont participé. Les résultats indiquent que l'évaluation a influencé leur apprentissage par le biais de plusieurs mécanismes différents. Certains d'entre eux nécessitaient une implication faible de la part de l'étudiant, comme l'identification des objectifs à atteindre (*feed-up*), l'apprentissage amélioré par les tests, la recherche d'informations après un examen. D'autres exigeaient une implication importante (par exemple, étudier pour les tests, sélectionner les évaluateurs pour obtenir une rétroaction de qualité, mettre à profit la rétroaction). L'implication des étudiants était modulée par leur perception des avantages et des inconvénients de s'impliquer, et de la crédibilité du système. Cette dernière était influencée par les objectifs-en-contexte des étudiants: devenir un bon médecin, contribuer à l'équipe soignante, réussir les épreuves d'évaluation.

Discussion : Notre système d'évaluation n'a pas réussi à impliquer suffisamment les étudiants que pour réaliser son potentiel. Nous abordons les défauts inhérents au système ainsi que les facteurs externes qui ont entravé l'implication des apprenants. Pour implanter efficacement un dispositif d'évaluation pour l'apprentissage, les concepteurs d'évaluations devraient optimiser les mécanismes qui sont faciles à contrôler et être prêts à s'investir dans un important travail de collaboration pour changer les cultures d'apprentissage.

Introduction

Assessment can influence as well as gauge learning.[1,2] Assessment can direct effective self-regulated learning behaviours during exam preparation,[3,4] reinforce prior learning through memory retrieval (test-enhanced learning),[5,6] and yield useful feedback information (catalytic effect).[2,7] Leaders in health professions education (HPE) have invoked the potential beneficial effects of assessment to call for curriculum designers to adopt assessment-for-learning practices.[8–10] Assessment-for-learning is "an approach in which the assessment process is inextricably embedded within the educational process, which is maximally information-rich, and which serves to steer and foster the learning of each individual student to the maximum of his/her ability."[8] This assessment approach involves both teachers and learners actively engaging with the assessment by reflecting on the information generated by assessments and using it in subsequent teaching and learning. However, HPE learners do not always use assessment to learn. Harrison et al. provided third-year medical students with digital feedback reports on their performance in an objective structured clinical examination.[11] Unfortunately, students who most needed feedback, accessed and engaged with it least.[11] A culture of assessments-as-hurdles directed students' attention to their scores rather than the feedback.[12] In contrast to traditional end-of-year assessments, programmatic assessment includes frequent assessments to provide on-going feedback. Although the assessments ultimately support pass-fail decisions, individual assessments do not constitute hurdles. Yet, the impact of programmatic assessment on learning has proved inconsistent.[13–16] Despite the *intended* low stakes of individual assessments, *perceived* stakes may lead to reduced or strategic engagement with the system, hampering its effectiveness.[14] Even when assessments have (and are perceived to have) no stakes, learners may deem the assessments meaningless or lacking in credibility, and not worthy of deep reflection; thus, unlikely to support learning.[17,18]

The purpose of our study was to contribute to a growing body of work investigating assessment-for-learning in HPE by designing, implementing, and evaluating an assessment system intended to support learning. Recognising that learner perceptions play a key role in assessment-for-learning, we focused our evaluation on learner perceptions of and experiences with the assessment system.

Methods

Context

Amid curricular reform, we implemented a mandatory course consisting solely of an assessment-for-learning system (details in Table 1) alongside clinical courses in the third year of a four-year undergraduate medical curriculum at McGill University. Following published recommendations,[19] the course was designed to 1) Communicate clear expectations of the program's objectives; 2) Elicit evidence of performance; 3) Provide detailed feedback several times a year; 4) Encourage students to take ownership of their learning and to study in a way that encourages retention and integration (both across clinical courses, and across the pre-clinical and clinical phases of the program); and, 5) Encourage retention of knowledge through testing. Progress tests have been widely used for these purposes.[20–22] End-of-clinical-rotation written exams were eliminated, with the hope that the progress test would encourage students to study in a deep fashion, integrating knowledge across their clerkship rotations. We also selected a workplace-based assessment tool, the Professional Mini-Evaluation Exercise (PMEX), that could be used in any discipline and that covered several competencies.[23] To maximise student ownership of learning and the perceived credibility of workplace-based feedback, students selected their raters.

Data collection

We conducted individual semi-structured and think-aloud interviews with students. We invited Year 3 and 4 students in the 2016-2017 academic year through class newsletters and student representatives. Participation was voluntary. Participants provided written consent before the interviews.

Two research assistants, with no involvement in the curriculum (to reduce the risk of socially-desirable responses), conducted the interviews using a semi-structured interview guide (Appendix 1). The interviewer showed students their latest results and asked them to think-aloud as they reviewed the materials. Finally, they showed participants a sample of 10 questions they had completed incorrectly on the previous progress test and asked them to think aloud as they answered these questions. Interviews lasted 60-90 minutes, were audio-recorded and transcribed verbatim, with identifying information removed. The Institutional Review Board of the Faculty of Medicine, McGill University, granted ethical approval (IRB # A01-E03-17A).

Table 1. Description of the two components of the assessment-for-learning intervention

	Progress test	Workplace-based assessment
Intended influences on learning	Pre-test effects: students engage in integrated and deep learning about all clinical disciplines and review basic sciences knowledge throughout the year Test-enhanced learning Feedback	Feedback
Content	Blueprinted to CanMEDS[49] with main focus on medical expert role, and to Medical Council of Canada clinical presentations (https://www.mcc.ca/objectives/expert/)	Professional behaviours associated with 6 of the 7 CanMEDS roles (professional, communicator, collaborator, health advocate, scholar, leader)
Tools and procedures	124 single-best answer multiple-choice questions Developed locally, some items translated from items shared by 3 other medical schools in the same province The tests were computer-based and administered in an invigilated environment in the medical school.	Professionalism Mini-Evaluation Exercise [23]: 22 rating-scale items 1 yes-no item regarding the occurrence of a critical event 1 open-ended item for comments To maximize the likelihood of feedback assimilation, we empowered students to select raters, hoping they would do so on the basis of raters' perceived credibility. To implicitly promote interprofessional collaboration, we allowed them to select assessors from any healthcare profession.
Frequency	Administered 3 times in year 3 (also once in year 2, and twice in year 4)	Elicited 6 or 12 times in year 3 (1 in each of the core clinical placements in the first year of implementation, and increased to 2 in each clinical placement in subsequent academic years)
Feedback	Feedback report sent to students contained: - Overall score and class average - Subscores for each clinical presentation, with numbers of questions and class averages - Subscores for non-medical expert roles, with numbers of questions and class averages	Completed forms were available for student review on an online platform
Stakes	Low Students who scored below the cut-score (60% of the 95th percentile score of their class) were required to attend a group teaching session where questions and answers were discussed Scores were not included on academic transcripts (pass-fail curriculum)	Low Students had to elicit minimum number of forms Critical incidents would have been reported to the course committee (none were reported during the course of the study) Neither comments nor scores were included on academic transcripts

Analysis

We conducted thematic analysis of the data, starting with line-by-line inductive open-coding (conducted by the two research assistants and the first author) followed by theory-informed[3-5,7,24] structuring of the codes, and data theming by VD with iterative feedback from the team as analysis proceeded.[25,26] The theories that informed analysis included Cilliers et al.'s model of the pre-test effects of assessment,[3,4] Sargeant et al.'s findings of reflection as "a link between receiving and using assessment feedback,"[24] Hattie and Timperley's concepts of feed-back (information about where the learner is), feed-up (information about where the learner should be going), and feed-forward (information about how the learner should proceed to get there),[7] and test-enhanced learning.[5]

Reflexivity statement

The authors hold strong beliefs about the power of assessment and in various ways seek to leverage it in their professional work, as assessment designers (VD, MW), clinical educators and curriculum designers (RC, SC), and as experts in assessment validity (MW, MY). While this could direct their gaze to the beneficial effects of assessment, team members have previously examined the gap between assessment design hopes and implementation disappointments.[27,28] The authors have diverse disciplinary backgrounds, i.e. medicine (VD, SC, RC), cognitive psychology (MY), and language assessment (MW), bringing diverse theoretical lenses and practical experience in assessment to the project. VD was the only author involved in the assessment system.

Results

Eleven students out of two classes of approximately 185 students agreed to participate, seven from Year 3, and 4 from Year 4. Figure 1 depicts a summary of findings. Table 2 provides a selection of illustrative quotes.

Table 2. Illustrative quotes. Participants (n = 11) were identified as P1 to P11.

The effects of assessment on learning depend on mechanisms that involve varying degrees of learner engagement	
Feed-up The assessment system signals program objectives	*That was a very busy rotation and there was a lot of time pressure and it was very good to hear some people recognize that, yes you are also behaving really professionally with the patients and you know, being attentive to their needs. And ya, so what it sort of did for me was that I kept doing that. You know I think if there had been complete silence, I might have given up more of that in the hopes of becoming more efficient. Because I was sort of getting the whole vibe of the place was efficiency, efficiency. (P7)*
Test-enhanced learning Learning occurs during the assessment itself	*I feel like there's umm this effect that we see the question so then at one point its gonna stick. (P11)*
Looking it up The experience of uncertainty during the assessment triggers a chase for answers right after the test (either through questioning peers or looking it up)	*I looked that up when I got home because I was like 'Oh I don't know this' and then I brought it up when we were having a group discussion. And that was an example of saw it on the test, looked it up at home, and like applied it in the clinical context and then looked like I knew something. (P8)*
Studying The assessment encourages and steers learning in preparation for it	*It did make me want to do more ongoing studying of subjects not necessarily related to the rotation I'm in. (…) I think there's something useful for continuous reviewing of material. (P7)*
Selecting raters who will provide quality feedback Students purposefully select raters with the expectation that these particular raters will provide helpful feedback	*(I selected assessors) who would be able to comment on my skills to some degree. They also had sat down with us at a midway point to give us feedback personally. So I knew that they had some knowledge of who I was, what skills and abilities and professionalism as well. (P10)*
Using feedback Students use – or could potentially use – the feedback	*I'm grateful for this (the progress test report), because in terms of information, this is one of the most useful information you can have about what you are doing and what you are not doing. I basically use this to see 'Oh I don't know about this, let me take a look at this.' (P6)*
	(The PMEX report is) clear (…), like its not super vague like medical expert or advocacy or something like that. It's simple things that you can work on. (P1)
Disengaging Students do not expend any time or effort in behaviours associated with learning for, during, or from the assessment, either from the outset or after initially doing so	*I don't feel that it pushes me to study and I'm happy that I don't have this stress in my life but at the same time I feel like I might need a bit more pushing to read more around the cases to prepare for LMCC's (licensing exam) next year. (P1)*
	I became less nervous (about selecting assessors). [Laughs] It was just ones of those things where it was a means to an end like I had to get it done to get a completion on my course. Like I didn't have to…like the first time I wanted to select people who had seen me in an encounter and really knew me and then the next times it was just like 'Who can fill this out?'. (P2)
Learner engagement depends on learner perceptions of the credibility of assessment and feedback, of the costs involved, and of the benefits involved	
Credibility Students gauge the trustworthiness of the assessments and/or the feedback reports	
Progress test	*I think the progress test is a very good thing, I think that they are an excellent way of getting us prepared for the LMCC (licensing exam) but just potentially like I don't know if the questions, and this is something I can't answer until I write the LMCC (licensing exam), but I don't know if they're very aligned with what we're gonna see on that final exam. (P1)* *I'm also concerned by the fact that I hear from my peers in the year ahead of me, like yah the average is always 60% and everyone just gets 60% and we've always gotten 60% and we always will get 60%, we're not improving as we gain more knowledge. I'm not sure why. (P3)*
Workplace-based component	*There's a few I think that wrote feedback or comments but most of them they just filled out the check boxes and that was it. And it took them a minute to do and it was over with so it seemed like it was just a means to an end. (P2)*
Benefits Students gauge the potential positive consequences of engaging in the assessment or feedback processes	*To get 64 on an exam without preparing for it, that covers the length and breath of your medical knowledge at the beginning when 60% is all you need to pass at the end, did not motivate me to study for the subsequent test. (P10)*
Costs Students gauge the potential negative consequences of engaging in the assessment or feedback processes	*So for example if you have 1 pediatrics exam that 121 questions, you need to know pediatrics in depth. But if it's overall, that's a maximum of 10 pediatrics questions that they are going to pick at random. So how much are you going to spend on that and then how much in depth do you really need to know? (P10)*
Learner perceptions of the utility of assessment and feedback depend on learner goals-in-context The term goals-in-context highlights that learner goals are influenced by the context in which they perceive themselves to be	
	At this stage in my training, I feel very uncomfortable asking someone to watch, let's say a full interaction, because I know they are very busy and I don't know that's something that I would get much out of. Like I feel really targeting 'This is a particular case that I feel very uncomfortable with, can you please help me with this?'. Like in a way it's almost like I'm asking them to come in for the things where I know I will struggle. So that's not necessarily the situations that I want to be evaluated on. (P7)

The effects of assessment-on-learning depend on mechanisms that involve varying degrees of learner engagement

Engaged students studied across rotations and actively sought high-quality feedback from supervisors. However, most participants did not engage - or rapidly disengaged - from the assessment-for-learning process. Many did not study for the progress tests, despite their concerns about the consequences, and collected workplace-based assessment forms solely to comply with requirements, asking whomever was available to complete them, oftentimes retrospectively. Nevertheless, the workplace-based component conveyed the importance of Physicianship for students (a concept, encompassing the professional and healer roles, that is a core feature of our curriculum[29]), communicating program objectives (so-called feed-up[7]), and serving as a counterpoint to the prevailing focus on efficiency in some hospital settings. The progress test appeared to generate test-enhanced learning and prompted students to look things up or discuss questions with peers.

Learner engagement depends on learner perceptions of the credibility of assessment and feedback, of the costs involved, and of the benefits involved

Credibility. Participant perceptions of the credibility of the assessment system varied. For the progress tests, participants gauged credibility on clinical relevance, alignment with the licensing exam, and item quality.

For the workplace-based component, participant credibility perceptions were based on rater issues such as variability or perceived careless completion of forms; form issues such as minimal space for comments; context issues including concern about the bias introduced by students self-selecting raters; and comparisons with other benchmarks such as self-assessment.

Benefits. Participants understood the intended purposes of the assessments. They perceived an additional – unintended – purpose of the progress tests: to prepare them for the licensing exam. However, the system fell short in their view, as it lacked credibility, it failed to detect progress, and/or its low stakes did not trigger extrinsic motivation.

Costs. The broad scope of the progress tests decreased the expected return-on-investment of studying.

For the workplace-based assessments, participants were concerned about burdening their busy supervisors, particularly those they perceived as good educators and who were more likely to be asked to complete assessments.

Learner perceptions of the usefulness of assessment and feedback depend on learner goals-in-context

Participants revealed three types of goals that influenced their appraisal of the assessment system: 1) contributing to the healthcare team; 2) becoming a good doctor; and 3) succeeding in (external) assessments.

Many participants wanted to contribute to the healthcare team's work and sought to limit the burden of workplace-based assessments on raters. The desire to become a good doctor manifested in the value placed on examinations' clinical relevance. Finally, success in assessments, especially external assessments, dominated the minds of some participants, who engaged/disengaged with the current assessment depending on whether they thought it would impact the outcomes of external assessments. Participants held multiple goals, some of which conflicted. For example, requesting direct observation when they lacked confidence in their skills could serve their goal of contributing to safe care, but impeded success in assessment if poor performance was documented. Therefore, they preferred undocumented verbal feedback to formal assessment for improvement.

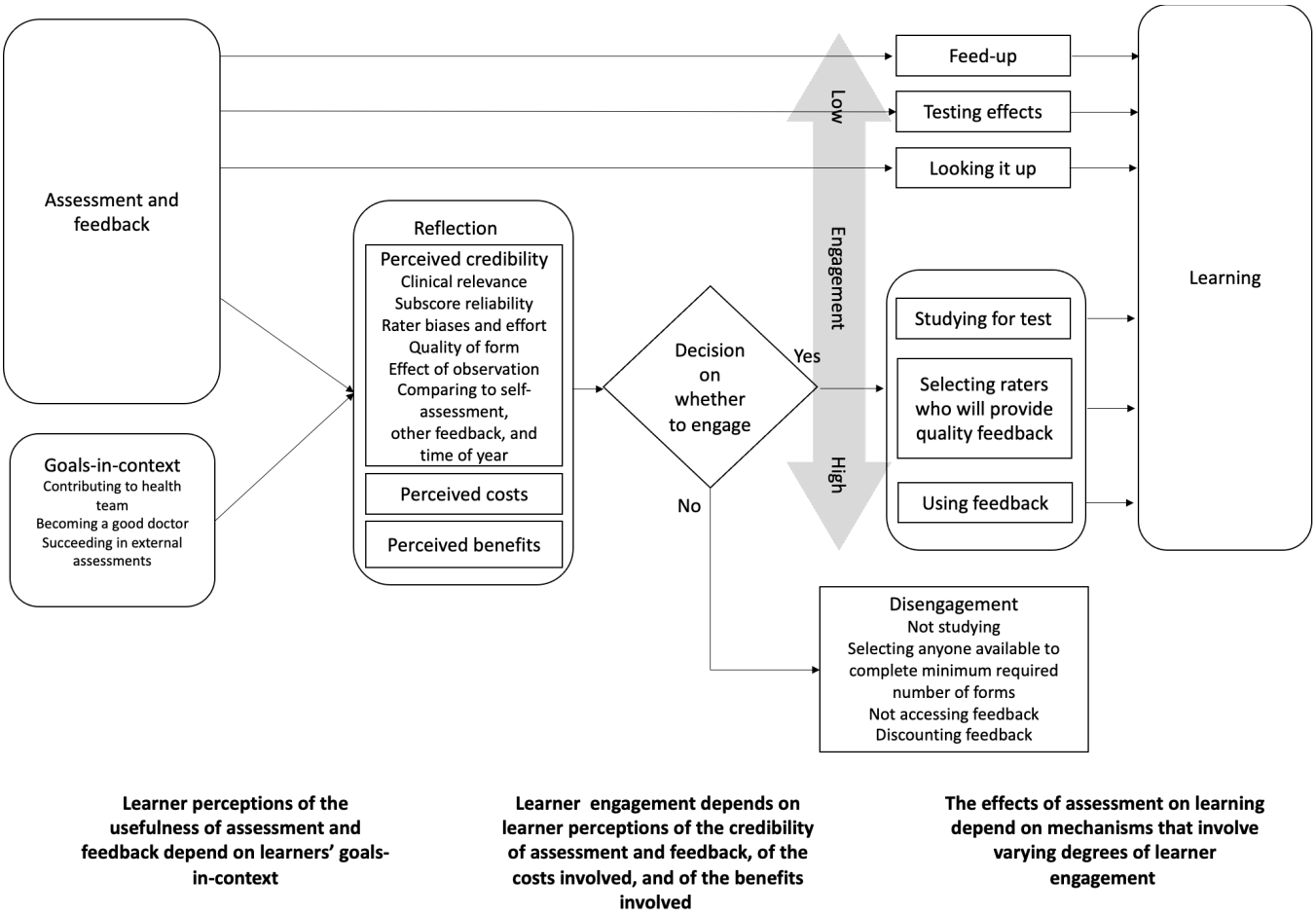

Figure 1. The effects of the assessment-for-learning intervention. The figure provides a visual depiction of the themes (in bold) we developed from our findings.

Discussion

Our assessment-for-learning intervention produced some desired effects when minimal student engagement sufficed. However, many students did not engage fully, limiting the intervention's effectiveness. This study has limitations. Our sample was a small convenience rather than purposive sample, which may limit the breadth and transferability of our findings. Our analysis is influenced by our backgrounds, beliefs, and knowledge.[30] While purposefully using existing theories illuminates interpretation, it may also selectively highlight and obscure different aspects of the data. Our use of cognitive theories and our transmission-oriented feedback lens[3-5,7,24] may have downplayed the interpersonal influences on student engagement.[31–33] Despite these limitations, our findings align with other work, lending them credibility. Others have previously found that learner engagement is critical[17,18,34,35] and hinges on a reflection process[24,36] that involves learners gauging the credibility of the assessment and of the feedback information[15,24,28,34–41] and the usefulness of engagement.[3,4,34,38,41–43] Context is known to influence learners' drive to use feedback.[12,17,38,41,44,45]

What then does this piece add to the conversation on assessment-for-learning? Our study distinguishes assessment-for-learning effects based on how much learner engagement is required. Effects that rely less on student engagement represent "low hanging fruit" for assessment designers, and we provide the following suggestions to maximise their potential. First, communicate the alignment of the assessment with program objectives, and when possible, with student goals, to leverage the "feed-up" process and steer self-regulated learning. Although our system aligned with program objectives, our participants were more concerned about alignment with the licensing exam. Anecdotally, student engagement increased after the first cohort experienced the licensing exam and we shared evidence that licensing exam and progress test scores correlated. Second, consider using open-ended questions on the progress test to maximise test-enhanced learning. In our context,

combining multiple-choice and open-ended key-feature questions would be required to maintain alignment with the licensing exam. Third, provide formal opportunities for students to discuss the test immediately after it has occurred to leverage post-examination learning from peers.

Effects that require more student engagement are inherently more complex and challenging to control and predict. To increase the likelihood of engagement, strategies should target students' perceptions of the benefits and costs of engaging. Involving students in the design of assessment may increase the perceived credibility of the assessment and alignment with student goals. This would likely require more co-design than what we did, which was to include student representatives on committees steering the curricular reform. However, aligning the assessment system with student goals is not straightforward. Students may hold conflicting goals and negotiate these conflicts in different ways. Furthermore, students' goals are influenced by external factors, such as licensing and residency-matching processes. Finally, aligning the assessment with student goals may have unintended negative effects on other processes involved in assessment-for-learning. For example, assessments may play a role in changing workplace culture and subsequently in changing future students' goals.

Decreasing the perceived costs of engaging is no easier. Students are reluctant to request assessment from busy clinicians, and clinicians may not all be trained in observation and feedback. Widespread faculty-development is onerous and the increase in clinical pressure is beyond the control of a given program. One avenue could be to recruit a cadre of dedicated clinicians with protected time for assessment.[46]

Finally, targeting the likelihood of engagement is not enough: engagement may also require specific skills to be effective. Recent work has highlighted the need to develop student feedback literacy or recipience,[47,48] which provides exciting avenues for research and educational practice.

Overall, increasing the likelihood and effectiveness of student engagement requires interventions at multiple levels, from involving students in co-designing the assessment and developing their feedback recipience, to training supervisors, to shifting work patterns in clinical settings, to influencing national licensing and residency-matching processes. In highlighting the challenges, we are not suggesting that assessment designers should give up on influencing these complex processes, rather, we are suggesting that the full potential of assessment-for-learning will require patience and effort to engage with all stakeholders.

Conflicts of Interest: Valérie Dory was involved in designing the assessment-for-learning system. The authors report no other conflicts of interest.

Funding: Class of 1984 gift to Undergraduate Medical Education, Faculty of Medicine, McGill University

Acknowledgements: The authors would like to acknowledge the participants who contributed their time and insights, the research assistants Jacqueline Safieh and Julia Kryluk who recruited and interviewed participants, transcribed the audio recordings, and contributed to the first cycle of coding, Dr Carlos Gomez-Garibello who provided feedback on the research protocol, and the McGill University MD class of 1984, spurred on by Dr Roland Grad, whose collective gift enabled the recruitment of research assistants for this study.

In Memoriam: We are saddened to be publishing this manuscript following the passing of our colleague Sylvia Cruess on 08/09/2022. VD, MW, and MY were privileged to have worked with, learned from, and just enjoyed the company of such a formidable and generous person. Sylvia was an exceptionally supportive member of the Canadian health sciences education community with a sharp sense of humor and a passion for continued learning. She will be greatly missed.

References

1. Kreiter C, Green J, Lenoch S, Saiki T. The overall impact of testing on medical student learning: quantitative estimation of consequential validity. *Adv Health Sci Educ*. 2013;18(4):835-844. https://doi.org/10.1007/s10459-012-9395-7
2. Norcini J, Anderson MB, Bollela V, et al. 2018 Consensus framework for good assessment. *Med Teach*. 2018;40(11):1102-1109. https://doi.org/10.1080/0142159X.2018.1500016
3. Cilliers FJ, Schuwirth LW, Herman N, Adendorff HJ, van der Vleuten CP. A model of the pre-assessment learning effects of summative assessment in medical education. *Adv Health Sci Educ*. 2012;17(1):39-53. https://doi.org/10.1007/s10459-011-9292-5
4. Cilliers F, Schuwirth L, Adendorff H, Herman N, Van der Vleuten C. The mechanism of impact of summative assessment on medical students' learning. *Adv Health Sci Educ*. 2010;15(5):695-715. https://doi.org/10.1007/s10459-010-9232-9
5. Larsen DP, Butler AC, Roediger HL. Test-enhanced learning in medical education. *Med Educ*. 2008;42(10):959-966. https://doi.org/10.1111/j.1365-2923.2008.03124.x
6. Larsen DP, Butler AC, Roediger Iii HL. Comparative effects of test-enhanced learning and self-explanation on long-term retention. *Med Educ*. 2013;47(7):674-682. https://doi.org/10.1111/medu.12141
7. Hattie J, Timperley H. The power of feedback. *Rev Educ Res*. 2007;77(1):81-112. https://doi.org/10.3102/003465430298487
8. Schuwirth LWT, Van der Vleuten CPM. Programmatic assessment: From assessment of learning to assessment for learning. *Med Teach*. 2011;33(6):478-485. https://doi.org/10.3109/0142159X.2011.565828

9. Schuwirth L. Making the horse drink: use of mini-CEX in an assessment for learning view. *Adv Health Sci Educ*. 2013;18:1–4. https://doi.org/10.1007/s10459-012-9423-7
10. Holmboe ES, Sherbino J, Long DM, Swing SR, Frank JR. The role of assessment in competency-based medical education. *Med Teach*. 2010;32(8):676-682. https://doi.org/10.3109/0142159X.2010.500704
11. Harrison CJ, Könings KD, Molyneux A, Schuwirth LWT, Wass V, Van der Vleuten CPM. Web-based feedback after summative assessment: how do students engage? *Med Educ*. 2013;47(7):734-744. https://doi.org/10.1111/medu.12209
12. Harrison C, Könings K, Schuwirth L, Wass V, van der Vleuten C. Barriers to the uptake and use of feedback in the context of summative assessment. *Adv Health Sci Educ*. 2015;20(1):229-245. https://doi.org/10.1007/s10459-014-9524-6
13. Driessen EW, van Tartwijk J, Govaerts M, Teunissen P, van der Vleuten CP. The use of programmatic assessment in the clinical workplace: a Maastricht case report. *Med Teach*. 2012;34(3):226-231. https://doi.org/10.3109/0142159x.2012.652242
14. Bok H, Teunissen P, Favier R, et al. Programmatic assessment of competency-based workplace learning: when theory meets practice. *BMC Med Educ*. 2013;13(1):123. https://doi.org/10.1186/1472-6920-13-123
15. Heeneman S, Oudkerk Pool A, Schuwirth LW, van der Vleuten CP, Driessen EW. The impact of programmatic assessment on student learning: theory versus practice. *Med Educ*. 2015;49(5):487-498. https://doi.org/10.1111/medu.12645
16. Bindal T, Wall D, Goodyear HM. Trainee doctors' views on workplace-based assessments: Are they just a tick box exercise? *Med Teach*. 2011;33(11):919-927. https://doi.org/10.3109/0142159X.2011.558140
17. Rees CE, Cleland JA, Dennis A, Kelly N, Mattick K, Monrouxe LV. Supervised learning events in the Foundation Programme: a UK-wide narrative interview study. *BMJ Open*. 2014;4(10). https://doi.org/10.1136/bmjopen-2014-005980
18. Barrett A, Galvin R, Scherpbier AJJA, Teunissen PW, O'Shaughnessy A, Horgan M. Is the learning value of workplace-based assessment being realised? A qualitative study of trainer and trainee perceptions and experiences. *Postgrad Med J*. 2017;93(1097):138-142. https://doi.org/10.1136/postgradmedj-2015-133917
19. Wiliam D, Thompson M. Integrating assessment with learning: What will it take to make it work? In: *The Future of Assessment: Shaping Teaching and Learning*. 1st ed. Lawrence Erlbaum Associates; 2008.
20. Freeman A, Van Der Vleuten C, Nouns Z, Ricketts C. Progress testing internationally. *Med Teach*. 2010;32(6):451-455. https://doi.org/10.3109/0142159X.2010.485231
21. van der Vleuten CPM, Verwijnen GM, Wijnen WHFW. Fifteen years of experience with progress testing in a problem-based learning curriculum. *Med Teach*. 1996;18(2):103-109. https://doi.org/10.3109/01421599609034142
22. Yielder J, Wearn A, Chen Y, et al. A qualitative exploration of student perceptions of the impact of progress tests on learning and emotional wellbeing. *BMC Med Educ*. 2017;17(1):148. https://doi.org/10.1186/s12909-017-0984-2
23. Cruess R, McIlroy JH, Cruess S, Ginsburg S, Steinert Y. The Professionalism Mini-Evaluation Exercise: A Preliminary Investigation. *Acad Med*. 2006;81(10):S74-S78. https://doi.org/10.1097/00001888-200610001-00019
24. Sargeant J, Mann K, Van der Vleuten C, Metsemakers J. Reflection: a link between receiving and using assessment feedback. *Adv Health Sci Educ*. 2009;14(3):399-410. https://doi.org/0.1007/s10459-008-9124-4
25. Saldaña J. *The coding manual for qualitative researchers*. 3rd ed. Sage; 2016.
26. Kiger ME, Varpio L. Thematic analysis of qualitative data: AMEE Guide No. 131. *Med Teach*. Published online May 1, 2020:1-9. https://doi.org/10.1080/0142159X.2020.1755030
27. Dory V, Gomez-Garibello C, Cruess R, Cruess S, Cummings BA, Young M. The challenges of detecting progress in generic competencies in the clinical setting. *Med Educ*. 2018;52(12):1259-1270. https://doi.org/10.1111/medu.13749
28. Ricci M, St-Onge C, Xiao J, Young M. Students as stakeholders in assessment: how students perceive the value of an assessment. *Perspect Med Educ*. 2018;7(6):352-361. https://doi.org/10.1007/s40037-018-0480-3
29. Boudreau JD, Cassell E, Fuks A. A healing curriculum. *Med Educ*. 2007;41(12):1193-1201. https://doi.org/10.1111/j.1365-2923.2007.02905.x
30. Young M, Ryan A. Postpositivism in health professions education scholarship. *Acad Med*. 2020;95(5):695-699. https://doi.org/10.1097/ACM.0000000000003089
31. Molloy E, Ajjawi R, Bearman M, Noble C, Rudland J, Ryan A. Challenging feedback myths: values, learner involvement and promoting effects beyond the immediate task. *Med Educ*. 2020;54(1):33-39. https://doi.org/10.1111/medu.13802
32. Ajjawi R, Regehr G. When I say ... feedback. *Med Educ*. 2019;53(7):652-654. https://doi.org/10.1111/medu.13746
33. Tavares W, Eppich W, Cheng A, et al. Learning conversations: an analysis of their theoretical roots and their manifestations of feedback and debriefing in medical education. *Acad Med*. 2020;95(7):1020-1025. https://doi.org/10.1097/acm.0000000000002932
34. Watling CJ, Kenyon CF, Zibrowski EM, et al. Rules of Engagement: residents' perceptions of the in-training evaluation process. *Acad Med*. 2008;83(10):S97-S100. https://doi.org/10.1097/ACM.0b013e318183e78c
35. Scarff CE, Bearman M, Chiavaroli N, Trumble S. Trainees' perspectives of assessment messages: a narrative systematic review. *Med Educ*. 2019;53(3):221-233. https://doi.org/10.1111/medu.13775
36. Eva K, Armson H, Holmboe E, et al. Factors influencing responsiveness to feedback: on the interplay between fear, confidence, and reasoning processes. *Adv Health Sci Educ*. 2012;17(1):15-26. https://doi.org/10.1007/s10459-011-9290-7
37. Gaunt A, Patel A, Rusius V, Royle TJ, Markham DH, Pawlikowska T. 'Playing the game': how do surgical trainees seek feedback using workplace-based assessment? *Med Educ*. 2017;51(9):953-962. https://doi.org/10.1111/medu.13380
38. Bowen L, Marshall M, Murdoch-Eaton D. Medical student perceptions of feedback and feedback behaviors within the context of the "educational alliance." *Acad Med*.

2017;92(9):1303-1312. https://doi.org/10.1097/acm.0000000000001632
39. Telio S, Regehr G, Ajjawi R. Feedback and the educational alliance: examining credibility judgements and their consequences. *Med Educ*. 2016;50(9):933-942. https://doi.org/10.1111/medu.13063
40. Moroz A, Horlick M, Mandalaywala N, T Stern D. Faculty feedback that begins with resident self-assessment: motivation is the key to success. *Med Educ*. 2018;52(3):314-323. https://doi.org/10.1111/medu.13484
41. Delva D, Sargeant J, Miller S, et al. Encouraging residents to seek feedback. *Med Teach*. 2013;35(12):e1625-31. https://doi.org/10.3109/0142159X.2013.806791
42. Bok HGJ, Teunissen PW, Spruijt A, et al. Clarifying students' feedback-seeking behaviour in clinical clerkships. *Med Educ*. 2013;47(3):282-291. https://doi.org/10.1111/medu.12054
43. Bing-You R, Hayes V, Palka T, Ford M, Trowbridge R. The art (and artifice) of seeking feedback: clerkship students' approaches to asking for feedback. *Acad Med*. 2018;93(8):1218-1226. https://doi.org/10.1097/acm.0000000000002256
44. Watling C, LaDonna KA, Lingard L, Voyer S, Hatala R. 'Sometimes the work just needs to be done': socio-cultural influences on direct observation in medical training. *Med Educ*. 2016;50(10):1054-1064. https://doi.org/10.1111/medu.13062
45. Ramani S, Könings KD, Mann KV, Pisarski EE, van der Vleuten CPM. About politeness, face, and feedback: exploring resident and faculty perceptions of how institutional feedback culture influences feedback practices. *Acad Med*. 2018;93(9):1348-1358. https://doi.org/10.1097/acm.0000000000002193
46. Voyer S, Cuncic C, Butler DL, MacNeil K, Watling C, Hatala R. Investigating conditions for meaningful feedback in the context of an evidence-based feedback programme. *Med Educ*. 2016;50(9):943-954. https://doi.org/10.1111/medu.13067
47. Molloy E, Boud D, Henderson M. Developing a learning-centred framework for feedback literacy. *Assess Eval High Educ*. 2020;45(4):527-540. https://doi.org/10.1080/02602938.2019.1667955
48. Winstone NE, Nash RA, Parker M, Rowntree J. Supporting learners' agentic engagement with feedback: a systematic review and a taxonomy of recipience processes. *Educ Psychol*. 2017;52(1):17-37. https://doi.org/10.1080/00461520.2016.1207538
49. Royal College of Physicians and Surgeons of Canada. *CanMEDS 2015: physician competency framework*. (Frank JR, Snell L, Sherbino J, eds.).; 2015.

Appendix A: Interview guide

Integrated assessment course

Here is an overview of the program objectives with Clerkship objectives in bold. To what extent do you think the Integrated Assessment course, that is the combination of the progress tests and PMEX forms, reflects the program and clerkship objectives?

PMEX

1. How do/did you go about requesting your first PMEX assessment?

   Prompts: How did you select the rater?
   Why did you ask him/her to assess you at that particular time?
   Was there anything else you were thinking about back then?

2. Were there any differences in how you went about it for other PMEX assessments?

   Prompt:
   - Were there differences by time of year? By discipline? By site?

3. Can you tell me how the forms were completed?

   Prompts:
   - Were they completed immediately after you ask it for?
   - Did you have the chance to discuss the assessment with the attending?
   - Were you able to review the forms?

4. Let's look at this specific PMEX form. I'd like you to tell me everything that's going through your mind as you read through it.

5. Can you go back to when you first received this feedback and tell me about it?

   Prompts:
   - How did you feel after reading this feedback?
   - What did you think about the feedback?
   - Did your thinking change in the subsequent days?

6. How did you use the feedback you received?

   Prompts:
   - How did the feedback help you to delineate your strengths and challenges?
   - Did you decide to improve on anything specific? If so, how did you go about working on that aspect?
   - How did you determine whether you had in fact improved on that aspect?
   - If you did not decide to improve on anything specific, how did you come to that decision?

7. How, if at all, is PMEX influencing/did PMEX influence your learning in Clerkship?

Progress test

8. Can you tell me about your experience with the progress test?

9. Let's look at this specific progress test report. I'd like you to tell me everything that's going through your mind as you read through it.

   Prompts:
   - How do you feel after reading this feedback?
   - What do you think about the feedback?

10. How, if at all, will/did you use this feedback?

    Prompts: How does the feedback help you to delineate your strengths and challenges?
    Will you decide to improve on anything specific? If not, please explain. If so, how will you go about working on that aspect?
    How will you determine whether you have in fact improved on that aspect?

11. How, if at all, is the progress test influencing your learning in Clerkship?

Learning in Clerkship

12. How do/did you determine what you are/were going to put effort into to learn in the clinical setting?

13. How do you determine what and how you are going to study?

14. Are there other ways in which you gauge how you are doing during Clerkship?

15. How are they different to PMEX and the progress test?