50TH ANNIVERSARY SERIES

Evolution or revolution to programmatic assessment: Considering unintended consequences of assessment change

A. Ryan[1], D. O'Mara[2] & M. Tweed[3]

Abstract

Assessment in the health professions is transforming. The widespread dominance of a reductionist measurement-based approach over the past 50 years is shifting towards a preference for more authentic assessment designed to promote and support learning. Assessment as a series of individual barriers, each to be surmounted, is being discarded in favour of systems of assessment designed to scaffold learner development and ensure sufficient opportunities for achievement. The intentions of these changes are to avoid the negative impacts of previous assessment approaches, such as strategic gaming, unhealthy competition and a predominance of book study, over immersion in clinical environments. However, unintended outcomes need to be considered when planning such transformative assessment change—both for those engaged in incremental evolutionary change and for those taking a more rapid or revolutionary approach. We explore three key features of programmatic assessment: longitudinal use of multiple assessment formats, a focus on assessment for learning and collation of data by attribute for decision making. We highlight the intended and possible unintended outcomes related to these features from the perspective of evolutionary and revolutionary approaches to change. We postulate that careful consideration of unintended outcomes is essential when planning significant assessment redesigns in health professional education. Anticipating unintended outcomes might also provide both the motivation and rationale to advance assessment practice into the next 50 years—particularly in the areas of enhancements in technology and collaborations across and between education providers.

Keywords: medical; health professions; education; assessment; design; programmatic

[1] Department of Medical Education, Melbourne Medical School, The University of Melbourne, Melbourne, Victoria, Australia
[2] Education Office, Sydney Medical School, Faculty of Medicine and Health, The University of Sydney, Sydney, NSW, Australia
[3] Department of Medicine, University of Otago, Wellington, New Zealand

Correspondence: Prof. Anna Ryan annatr@unimelb.edu.au

Introduction

Assessment in health professional education is transforming. The overriding dominance of a reductionist measurement-based approach is being challenged by more wholistic systems of assessment, authentic to practice and designed to support learning (Norcini et al., 2018; Schuwirth & van der Vleuten, 2020). The forces for change are diverse and compelling: the imperative for patient safety, increasing recognition of the importance of communication skills and professionalism, demonstrating quality assurance and also reducing unhealthy and competitive health professional cultures and strategic gaming of assessment tasks. Programmatic assessment (PA) (van der Vleuten & Schuwirth, 2005) prioritises the use of assessment to drive and support learning and emphasises the value of longitudinal assessment data collation by attributes for feedback and decision making.

While initial PA implementations primarily involved small cohorts or new schools (Heeneman et al., 2015; Schut et al., 2020), more recent implementations include larger, established schools (Ryan & Judd, 2022; Tait & Kulasegaram, 2022), with descriptions of the courses and approaches taken outlined in a recent review (Torre et al., 2021). Some have implemented incremental changes (an evolutionary approach, e.g., Ryan & Judd, 2022), whilst the alternative is wholesale program transformations (a revolutionary approach, e.g., Wilkinson & Tweed, 2018). While the rationale for change is persuasive and the intended aims are clear, the complexity of context and the many forces at play mean that PA's intended outcomes are not guaranteed and require assessment designers to be alert to the potential for unintended consequences.

Framework

Consideration of unintended consequences was introduced to the policy evaluation framework literature by Robert Merton (1936). The public policy literature has a long tradition of drawing on existing evidence to develop and evaluate policy to address complex social problems. Interventions are seen as dynamic, realised through the responses of multiple stakeholders embedded in complex social systems and influenced by multiple concurrent interventions (Pawson et al., 2011). In this context, it is not possible to know with certainty all likely outcomes of an intervention, nor to create a true experimental setting. Hence, deliberate consideration of potential unintended consequences is a key step in policy development, along with a continuous cycle of evaluation and improvement (Hadjar & Becker, 2009).

The parallels between policy interventions for complex social problems and systematic assessment change in health professional education are noteworthy. However, in contrast to public policy, it is not possible to introduce changes in a medical program as a "test site" because all university policy mandates that enrolled students within an institution must be given equivalent opportunities for learning. As such, we focus on three key elements of PA, drawing on the existing literature and our own experiences to highlight intended and possible unintended consequences of assessment change (Table 1).

Table 1

The Intended and Unintended Consequences of Three Key Elements of PA

Key Elements of PA	Intended Consequences	Unintended Consequences/Risk
Longitudinal use of multiple assessment data points	Decreased focus on a single high-stakes assessment Increased student engagement Reduced failure to fail	Excessive assessment workload Increased student anxiety Unlimited opportunities to meet the standard
Focus on authentic assessment for learning	Supports individualised approach More timely feedback relevant to clinical practice	Learning potential not reached due to other factors Complexity of feedback overwhelms the learner
Collation of data by attribute for decision making	More authentic assessment data points Improves rigour of progression decisions	Unmanageable data complexity Devaluing of individual assessments

Discussion

While lauded for their efficiency, the potentially detrimental impact of single **end of year high stakes assessments, or barriers**, has long been argued in the literature (O'Rourke et al., 2010). Although several models support longitudinal use of multiple assessment formats (Hoang & Lau, 2018; Pearce & Tavares, 2021; Schuwirth & van der Vleuten, 2011), introducing more frequent assessments over multiple data points necessarily increases workload for staff and students. Students can feel that every low-stakes assessment is like a "barrier" (Roberts et al., 2022) despite assurances otherwise. If introduced too quickly, through a revolutionary change, and without the appropriate increase in staff numbers and expertise (e.g., psychometrics), resources (e.g., eportfolio) and support for students, this change will be viewed as unnecessarily burdensome for all. Evolutionary changes have different challenges, with the need to keep traditional assessment barriers during the transition to a full PA system (Ryan & Judd, 2022).

Longitudinal assessment also carries with it the expectation of **increased student engagement** throughout the entire course, providing less opportunities for students to not attend learning opportunities and/or engage in end-of-year cramming. Research has indicated that longitudinal assessment can be associated with continuous pressure and anxiety (Bok et al., 2013; Heeneman et al., 2015; Schut et al., 2018; Schut et al., 2021). We postulate that student anxiety may also be influenced by uncertainty in revolutionary change, particularly if modifications are required due to unintended outcomes. Fear of missing out compared to other cohorts may be heightened in more evolutionary policy changes.

Longitudinal clinical assessment is promoted as a solution to the problem of **failure to fail** (Mak-van der Vossen, 2019). With each assessment being considered as a low stakes

data point, students who have not demonstrated the required level of competency at the same time as their peers are supported with further learning and/or additional assessment opportunities. However, a possible unintended consequence is that students meet the expected standard due to constant repetition with chance of higher performance (Albon, 2001) rather than the assessments being a valid reflection of increased competency. This risk can be reduced in a totally programmatic system where new data is aggregated with prior information rather than viewed as an independent hurdle.

The **focus on assessment for learning** is a compelling rationale for a PA approach and supported by early implementations (Bierer et al., 2015; Heeneman et al., 2015). The ideal outcome is that collation of longitudinal data across format would facilitate an individualised approach to learning, driving each student to focus on areas for improvement (Ross et al., 2021; Schuwirth & van der Vleuten, 2019). However, there are a number of reasons why assessment data is not used for its learning potential in a PA system, such as inappropriate staffing levels and expertise, different rates of achievement and aiming for the minimum standard rather than excellence (Wise & DeMars, 2005). Implementing supportive programs with a focus on growth mindset and mastery learning and mandating a requisite number of workplace-based assessments from "certified assessors" are potential mitigations.

There is value and complexity when dealing with **narrative feedback** (Schuwirth et al., 2017). Collation of narrative information presents a greater challenge and can rely heavily on staff and technology (Torre et al., 2021; Wilkinson & Tweed, 2018). Giving students high volumes of written feedback can be counterproductive (Glover & Brown, 2006), and aggregation of data by competency frameworks assumes this will be meaningful for students. The large and diverse PA data set must be effectively managed and collated. Helping students to understand the potential value of assessments is a core component of assessment literacy work whether change has been introduced in an evolutionary or revolutionary manner—highlighting the importance of student and staff co-design and supportive interpersonal relationships (Schut et al., 2021) and fit for purpose data systems (Ryan & Judd, 2022).

Another pedagogical strength of PA is **collation of data by attribute to make progress decisions**. Reductionist approaches to handling assessment data are limited (Schuwirth & van der Vleuten, 2020). Competency frameworks used in progression decisions have been developed as being authentic to practice (Bandiera et al., 2006; Graham et al., 2007; Medical Council of New Zealand, 2014). Theoretically, within a PA system, the data available to progress decision making panels will be content-rich narrative information (de Jong et al., 2022; Torre et al., 2021). If the data is well presented, a student who is clearly satisfactory on all assessment tasks will not need extensive discussion (Pearce et al., 2021), However with multiple data points, there are more students who do not meet the expected standard compared to the previous reductionist barrier-type approach. If the data is not effectively prepared for the purpose of decision making, then potential unintended

consequences occur: decision paralysis, an over reliance on the solution of computer-based algorithms, decisions based on selected information or reverting to reductionist decision making. The preparation and collation of data points may require staff with specialist psychometric and/or technology skills to ensure sufficient robustness and to make maximum use of both the qualitative and quantitative information. Although the evolutionary approach may allow for progressive development of data collation and decision making, two of our medical schools have encountered challenges in supporting staff to make wholistic decisions regarding progression, highlighting the importance of new decision making and staff support processes (Schut et al., 2021; Tweed et al., 2013; Tweed & Wilkinson, 2019) and the imperative of supportive leadership (Schut et al., 2021).

A significant proposed benefit of PA is that decisions are separate from assessment events and are based on **multiple data points**, collated over time to allow students to demonstrate development, all within a framework that is authentic (Wilkinson & Tweed, 2018). However, this kind of data collation carries the risk that staff will perceive individual subjects are devalued and diluted by compensation and that students will undervalue particular subjects/assessments, leading to disengagement (Pugh & Zhao, 2003). The intended use of each piece of assessment data is another important consideration. One of the authors noted, in their context, that narrative comments from learning advisors were limited when the aim was to include them in progress decisions but became more valued for learning when not used as a basis for the ultimate pass/fail decision. High-stakes exams build trust in a system, and conversely, lack of high stakes exams can undermine trust (Lyons, 2017; Richardson, 2022). Policies that introduce a range of lower stake assessments have the potential unintended consequence of stakeholders perceiving a compromise of standards. The importance of stakeholder consultation has been emphasised in the literature (Schut et al., 2021), and in one of our contexts, the development of a written assessment strategy has been helpful in sharing the intent of assessment change (Ryan & Judd, 2022). In an evolutionary change, there is more time and flexibility to get the systems in place and engage stakeholders (Kang et al., 2022), yet it carries the risk of different systems of progression decision making running in parallel, creating a perceived risk of loss of rigour. In a revolutionary model, the urgency for change creates additional pressures but has the advantage that there is a single system in place.

Conclusion

We use this manuscript in the ANZAHPE 50th anniversary series to reflect on evolutionary and revolutionary approaches to changing from more traditional to programmatic systems of assessment and also to look for what might occur in a better future state. The limitations of traditional assessment approaches are well recognised, yet systems approaches, such as PA, are not without risk. The existing evidence base provides

helpful guidance (Henneman et al., 2021; Torre et al., 2021), yet the wide variation of implementation contexts as well as the type of change implementation (evolutionary or revolutionary) make anticipation of unintended consequences challenging.

Evolutionary change approaches have the advantage of providing time for stakeholders to become accustomed to new approaches but also prolong the time for antagonists to delay change. In traditional university cultures, it might also provide the time required to implement required institutional policy change and to seek approval for the appropriate resources to support the educational change. Evolutionary change has the potential for development to occur gradually alongside business as usual, potentially saving costs, yet can overburden existing staff. It also provides the opportunity for small scale and pilot innovations that can be used to test approaches, solve early issues and build stakeholder engagement. Yet, the presence of variable approaches within the one course creates the risk of uncertainty and confusion. Particularly risky with this approach is the timepoint for transition to the new approach. If this is done too early, then there is potential for incorrect progression decisions due to less rigour and/or inadequate faculty development or resources and/or running two systems simultaneously for longer than anticipated.

Revolutionary change approaches mandate change at a certain timepoint and can be perceived as a simpler approach and easier for managing cohorts. These approaches are usually planned without the consideration of unintended consequences and often underestimate required resourcing (staff and financial) because high perceived cost might function as a barrier to implementing change. If not appropriately resourced, or if it occurs in parallel with other revolutions, such as a pandemic or major curriculum change, staff fatigue is likely in some cases, leading to resentment and antagonism. Revolutions carry the benefit of consistency and alignment across all years of the program or for a commencing cohort but often have an inflexible "go live" day, which can create major challenges if all processes are not in place at that time.

While PA implementations can and have been reported to result in positive change, excessive workload, increased resourcing requirements, perceptions of lower standards, excessive data complexity, poor decision making, persistent failure to fail, continued strategic gaming and increased student stress or demotivation are all potential unintended consequences. A better future state would have mitigations for these known, but also emerging, unintended consequences.

Clear articulation of assessment strategy and rationale for change, staff training, development of supportive policy and procedures for a programmatic approach, and careful attention to stakeholder engagement are all strategies that have potential to address and mitigate some of those unintended consequences. Appropriate resourcing, especially appropriate data management using fit for purpose systems, a focus on student assessment literacy as well as supportive relationships and student codesign are also essential when implementing significant assessment transformations.

The imperative to improve assessment and support mastery learning is currently having a major impact on systems of assessment in health professional education. Whether this approach is transitory or results in long-term transformative change remains to be seen and is likely heavily dependent on avoidance of unintended consequences. We believe the potential for programmatic assessment approaches to reach their full promise over the coming decades is heavily reliant on two key advances in the areas of technology and collaboration.

Two current significant unintended consequences of a PA system are the volume and complexity of feedback data overwhelming the learner and of progression-informing data overwhelming staff. Any problem related to data volume and complexity is likely to have future technology-based solutions. Over the next 50 years, these will likely relate to assessment data acquisition, collation, synthesis, accessibility and presentation. This will allow a move from any form of tick-box exercise or review of disparate events for the learner to collated evidence of a rich learning journey. Progression decisions could be supported by artificial intelligence (AI), for example, providing a summary of the rich narrative feedback for multiple workplace-based assessments over time. In the Australian and New Zealand context, eportfolio systems developed for internship or specialty training could be adapted by medical schools and could support the implementation of PA in more schools. However, more collaboration between schools will be required to align assessment systems and might well be resisted as undermining medical school autonomy and individuality.

Nevertheless, there are many opportunities for collaboration with respect to PA developments, and these will become more important in the future. We anticipate this will occur across undergraduate education, through post-graduate training to continued professional development—across countries and across healthcare and other professions. As healthcare education resources become more constrained, collaboration to share and develop resources will occur from the level of assessment item development through to system governance. The technology advances mentioned will support these many levels of collaboration.

The risk of not embracing the imperative to collaborate is that we default to systems of education, training and assessment that require little in the way of resources, are reliant on an internet full of available material and a single national high stakes examination outsourced to the most efficient provider. If Australian and New Zealand national medical knowledge examinations were introduced, there could be positive and negative unintended consequences. On the positive side, medical schools could focus on clinical and workplace-based assessment, and assuming staff resources were appropriate, the quality of assessment experience and the capabilities of graduates may increase. National examination candidate numbers could allow for computer adaptive testing to make assessment more efficient and individualised. On the negative side, a high stakes national exam would have the unintended consequence of driving a content-based curriculum

focus, reinforce gaming, disadvantage those from groups underrepresented in medicine and decrease the diversity of participants in medical education in our two countries.

The public policy literature highlights the importance of consideration of potential unintended consequences alongside continuous cycles of evaluation and improvement throughout an intervention to detect and respond to unexpected and undesired consequences as they arise. We have highlighted challenges arising from the literature and our PA implementations with the aim of helping others to anticipate what might lie in store if they were to follow a similar path. Cognisance of the potential unintended consequences reduces the risk of entering into an evolutionary or revolutionary change to a system of programmatic assessment and replacing one range of problems from the past 50 years with a different set of problems for the next 50 years. We propose that advance consideration of unintended outcomes is essential and, in this context, highlights the importance of collaborating and harnessing technology to support transformative, meaningful and sustainable change in assessment practice.

Conflicts of interest and funding

The authors have no conflicts of interest or funding sources to declare.

References

Albon, R. (2001). Examine them until they pass! *Agenda: A Journal of Policy Analysis and Reform*, *8*(2), 185–191. https://www.jstor.org/stable/43199162

Bandiera, G., Sherbino, J., & Frank, J. R. (Eds.). (2006). *The CanMEDS assessment tools handbook: An introductory guide to assessment methods for the CanMEDS competencies*. The Royal College of Physicians and Surgeons of Canada.

Bierer, S. B., Dannefer, E. F., & Tetzlaff, J. E. (2015). Time to loosen the apron strings: Cohort-based evaluation of a learner-driven remediation model at one medical school. *Journal of General Internal Medicine*, *30*(9), 1339–1343. https://doi.org/10.1007/s11606-015-3343-1

Bok, H. G., Teunissen, P. W., Favier, R. P., Rietbroek, N. J., Theyse, L. F., Brommer, H., Haarhuis, J. C. M., van Beukelen, P., van der Vleuten, C. P., M., & Jaarsma, D. A. (2013). Programmatic assessment of competency-based workplace learning: When theory meets practice. *BMC Medical Education*, *13*(1), Article 123. https://doi.org/10.1186/1472-6920-13-123

de Jong, L. H., Bok, H. G., Schellekens, L. H., Kremer, W. D., Jonker, F. H., & van der Vleuten, C. P. (2022). Shaping the right conditions in programmatic assessment: How quality of narrative information affects the quality of high-stakes decision-making. *BMC Medical Education*, *22*(1), Article 409. https://doi.org/10.1186/s12909-022-03257-2

Glover, C., & Brown, E. (2006). Written feedback for students: Too much, too detailed or too incomprehensible to be effective? *Bioscience Education*, *7*(1), 1–16. https://doi.org/10.3108/beej.2006.07000004

Graham, I. S., Gleason, A. J., Keogh, G. W., Paltridge, D., Rogers, I. R., Walton, M., De Paola, C., Singh, J., & McGrath, B. P. (2007). Australian curriculum framework for junior doctors. *Medical Journal of Australia*, *186*(S7), S14–S19. https://doi.org/10.5694/j.1326-5377.2007.tb00959.x

Hadjar, A., & Becker, R. (2009). Educational expansion: Expected and unexpected consequences. In A. Hadjar & R. Becker (Eds.), *Expected and unexpected consequences of the educational expansion in Europe and the US* (pp. 9–23). Haupt Berne.

Heeneman, S., Oudkerk Pool, A., Schuwirth, L. W., van der Vleuten, C. P., & Driessen, E. W. (2015). The impact of programmatic assessment on student learning: Theory versus practice. *Medical Education*, *49*(5), 487–498. https://doi.org/10.1111/medu.12645

Hoang, N. S., & Lau, J. N. (2018). A call for mixed methods in competency-based medical education: How we can prevent the overfitting of curriculum and assessment. *Academic Medicine*, *93*(7), 996–1001. https://doi.org/10.1097/ACM.0000000000002205

Kang, S. P., Chen, Y., Svihla, V., Gallup, A., Ferris, K., & Datye, A. K. (2022). Guiding change in higher education: An emergent, iterative application of Kotter's change model. *Studies in Higher Education*, *47*(2), 270–289. https://doi.org/10.1080/03075079.2020.1741540

Lyons, P. (2017, December 26). Peter Lyons: NCEA teaching our kids they don't need to try too hard. *NZ Herald*. https://www.nzherald.co.nz/nz/peter-lyons-ncea-teaching-our-kids-they-dont-need-to-try-too-hard/R63AFSZ6MK346OCPGWJXIO47CE/

Mak-van der Vossen, M. (2019). "Failure to fail": The teacher's dilemma revisited. *Medical Education*, *53*(2), 108–110. https://doi.org/10.1111/medu.13772

Medical Council of New Zealand. (2014, February). *New Zealand curriculum framework for prevocational medical training*. https://www.mcnz.org.nz/assets/Forms/4b7ce95390/New-Zealand-Curriculum-Framework.pdf

Merton R. K. (1936). The unanticipated consequences of purposive social action. *American Sociological Review*, *1*(6), 894–904. https://doi.org/10.2307/2084615

Norcini, J., Anderson, M. B., Bollela, V., Burch, V., Costa, M. J., Duvivier, R., Hays, R., Palacios Mackay, M. F., Roberts, T., & Swanson, D. (2018). 2018 consensus framework for good assessment. *Medical Teacher*, *40*(11), 1102–1109. https://doi.org/10.1080/0142159X.2018.1500016

O'Rourke, M., Hammond, S., O'Flynn, S., & Boylan, G. (2010). The Medical Student Stress Profile: A tool for stress audit in medical training. *Medical Education*, *44*(10), 1027–1037. https://doi.org/10.1111/j.1365-2923.2010.03734.x

Pawson, R., Wong, G., & Owen, L. (2011). Known knowns, known unknowns, unknown unknowns: The predicament of evidence-based policy. *American Journal of Evaluation*, *32*(4), 518–546. https://doi.org/10.1177/1098214011403831

Pearce, J., Reid, K., Chiavaroli, N., & Hyam, D. (2021). Incorporating aspects of programmatic assessment into examinations: Aggregating rich information to inform decision-making. *Medical Teacher*, *43*(5), 567–574. https://doi.org/10.1080/0142159X.2021.1878122

Pearce, J., & Tavares, W. (2021). A philosophical history of programmatic assessment: Tracing shifting configurations. *Advances in Health Sciences Education*, *26*(4), 1291–1310. https://doi.org/10.1007/s10459-021-10050-1

Pugh, K. J., & Zhao, Y. (2003). Stories of teacher alienation: A look at the unintended consequences of efforts to empower teachers. *Teaching and Teacher Education*, *19*(2), 187–201. https://doi.org/10.1016/S0742-051X(02)00103-8

Richardson, M. (2022). *Rebuilding public confidence in educational assessment*. UCL Press.

Roberts, C., Khanna, P., Bleasel, J., Lane, S., Burgess, A., Charles, K., Howard, R., O'Mara, D., Haq, I., & Rutzou, T. (2022). Student perspectives on programmatic assessment in a large medical programme: A critical realist analysis. *Medical Education*, *56*(9), 901–914. https://doi.org/10.1111/medu.14807

Ross, S., Hauer, K. E., Wycliffe-Jones, K., Hall, A. K., Molgaard, L., Richardson, D., Oswald, A., Bhanji, F., & ICBME Collaborators. (2021). Key considerations in planning and designing programmatic assessment in competency-based medical education. *Medical Teacher*, *43*(7), 758–764. https://doi.org/10.1080/0142159X.2021.1925099

Ryan, A., & Judd, T. (2022). From traditional to programmatic assessment in three (not so) easy steps. *Education Sciences*, *12*(7), 487. https://doi.org/10.3390/educsci12070487

Schut, S., Driessen, E., Van Tartwijk, J., van der Vleuten, C., & Heeneman, S. (2018). Stakes in the eye of the beholder: An international study of learners' perceptions within programmatic assessment. *Medical Education*, 52(6), 654–663. https://doi.org/10.1111/medu.13532

Schut, S., Heeneman, S., Bierer, B., Driessen, E., van Tartwijk, J., & van der Vleuten, C. (2020). Between trust and control: Teachers' assessment conceptualisations within programmatic assessment. *Medical Education*, *54*(6), 528–537. https://doi.org/10.1111/medu.14075

Schut, S., Maggio, L. A., Heeneman, S., van Tartwijk, J., van der Vleuten, C., & Driessen, E. (2021). Where the rubber meets the road: An integrative review of programmatic assessment in health care professions education. *Perspectives on Medical Education*, *10*(1), 6–13. https://doi.org/10.1007/s40037-020-00625-w

Schuwirth, L. W., & van der Vleuten, C. P. (2011). Programmatic assessment: From assessment of learning to assessment for learning. *Medical Teacher*, *33*(6), 478–485. https://doi.org/10.3109/0142159X.2011.565828

Schuwirth, L. W., & van Der Vleuten, C. P. (2019). Current assessment in medical education: Programmatic assessment. *Journal of Applied Testing Technology*, *20*(S2), 2–10.

Schuwirth, L. W., & van der Vleuten, C. P. (2020). A history of assessment in medical education. *Advances in Health Sciences Education*, *25*(5), 1045–1056. https://doi.org/10.1007/s10459-020-10003-0

Schuwirth, L. W., van der Vleuten, C., & Durning, S. J. (2017). What programmatic assessment in medical education can learn from healthcare. *Perspectives on Medical Education*, *6*(4), 211–215. https://doi.org/10.1007/s40037-017-0345-1

Tait, G. R., & Kulasegaram, K. M. (2022). Assessment for learning: The University of Toronto Temerty Faculty of Medicine MD program experience. *Education Sciences*, *12*(4), 249. https://doi.org/10.3390/educsci12040249

Torre, D., Rice, N. E., Ryan, A., Bok, H., Dawson, L. J., Bierer, B., Wilkinson, T. J., Tait, G. R., Laughlin, T., Veerapen, K., Heeneman, S., Freeman, A., & van der Vleuten, C. (2021). Ottawa 2020 consensus statements for programmatic assessment—2. Implementation and practice. *Medical Teacher*, *43*(10), 1149–1160. https://doi.org/10.1080/0142159X.2021.1956681

Tweed, M. J., Thompson-Fawcett, M., & Wilkinson, T. J. (2013). Decision-making bias in assessment: The effect of aggregating objective information and anecdote. *Medical Teacher*, *35*(10), 832–837. https://doi.org/10.3109/0142159X.2013.803062

Tweed, M., & Wilkinson, T. (2019). Student progress decision-making in programmatic assessment: Can we extrapolate from clinical decision-making and jury decision-making? *BMC Medical Education*, *19*(1), Article 176. https://doi.org/10.1186/s12909-019-1583-1

van der Vleuten, C. P., & Schuwirth, L. W. (2005). Assessing professional competence: From methods to programmes. *Medical Education*, *39*(3), 309–317. https://doi.org/10.1111/j.1365-2929.2005.02094.x

Wilkinson, T. J., & Tweed, M. J. (2018). Deconstructing programmatic assessment. *Advances in Medical Education and Practice*, *2018*(9), 191–197. http://doi.org/10.2147/AMEP.S144449

Wise, S. L., & DeMars, C. E. (2005). Low examinee effort in low-stakes assessment: Problems and potential solutions. *Educational Assessment*, *10*(1), 1–17. https://doi.org/10.1207/s15326977ea1001_1