

Learning Analytics in Medical Education Assessment: The Past, the Present, and the Future

Teresa Chan, MD, FRCPC, MHPE, Stefanie Sebok-Syer, PhD, Brent Thoma, MD, MA, MSc, FRCPC, Alyssa Wise, PhD, Jonathan Sherbino, MD, MEd, FRCPC, and Martin Pusic, MD, FRCPC, PhD

ABSTRACT

With the implementation of competency-based medical education (CBME) in emergency medicine, residency programs will amass substantial amounts of qualitative and quantitative data about trainees' performances. This increased volume of data will challenge traditional processes for assessing trainees and remediating training deficiencies. At the intersection of trainee performance data and statistical modeling lies the field of medical learning analytics. At a local training program level, learning analytics has the potential to assist program directors and competency committees with interpreting assessment data to inform decision making. On a broader level, learning analytics can be used to explore system questions and identify problems that may impact our educational programs. Scholars outside of health professions education have been exploring the use of learning analytics for years and their theories and applications have the potential to inform our implementation of CBME. The purpose of this review is to characterize the methodologies of learning analytics and explore their potential to guide new forms of assessment within medical education.

A CASE OF A CLINICAL COMPETENCY COMMITTEE (CCC) FILE REVIEW

It was time for the CCC committee meeting and Dr. Zainab Hussain was not looking forward to file preparation. The data for each resident's file were presented in a spreadsheet, which was difficult to manipulate and hard to interpret. And yet, these were the data that her committee was supposed to use to guide their recommendations for annual progress review of the residents' milestone achievements. She sighed, remembering that it was her request for more robust data that led her program director and department chair to nominate her as the committee chair. In contrast to 5 years ago when decisions about

resident promotion were made with minimal information, they now had a lot of data points, but Zainab was not sure how to organize and analyze the data for effective interpretation. After all, they had to report residents' progress based on the national benchmarks, and she didn't want them to fall behind on reporting.

THE (BRIEF) HISTORY OF LEARNING ANALYTICS IN MEDICAL EDUCATION

Within the past decade, the field of analytics has exploded, and in medicine this is evidenced by the exponential growth in literature (see Figure 1). Many

From the Faculty of Health Science (JS), Division of Emergency Medicine, Department of Medicine, McMaster University and the McMaster program for Education Research, Innovation, and Theory (MERIT) (TC, JS), Hamilton, Ontario, Canada; the Centre for Education Research & Innovation, Schulich School of Medicine and Dentistry (SSS), Department of Emergency Medicine, University of Saskatchewan (BT), Saskatoon, Saskatchewan, Canada; the Steinhardt School of Culture, Education, and Human Development, New York University (AW), New York, NY; and the Department of Emergency Medicine, NYU School of Medicine (MP), New York, NY.

Received January 29, 2018; accepted January 30, 2018.

Dr. Chan's work on this paper has been supported by the W. Watson Buchanan Clinician Educator McMaster Department of Medicine Internal Career Award.

Dr. Martin Pusic would like to declare that he is a codeveloper of the ImageSim program. The other authors have no potential conflicts to disclose.

Supervising Editor: Susan Promes, MD, FACEP.

Address for correspondence and reprints: Dr. Teresa Chan; e-mail: teresa.chan@medportal.ca.

AEM EDUCATION AND TRAINING 2018;2:178-187

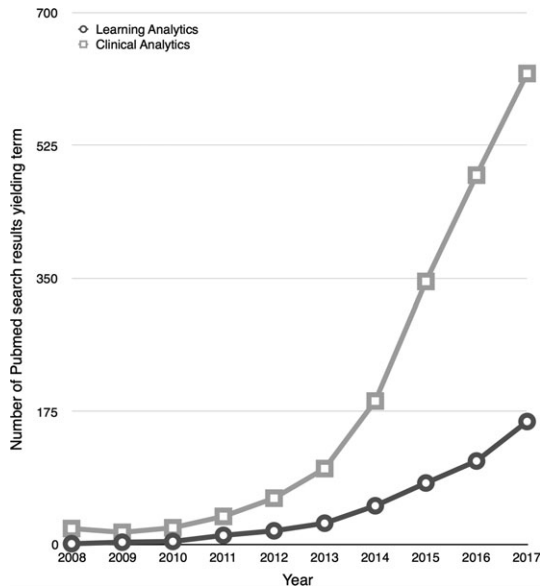


Figure 1. Number of PubMed citations over the past decade listing both the terms “clinical analytics” and “learning analytics.”

industries such as finance, sports, and security have benefitted from using data analytics; however, medicine has been slower to embrace these methods. Within medicine, the attention on analytics to improve education and training is less than that of clinical purposes, such as quality improvement. In short, the age of data analytics is upon us and medical education is lingering behind the times.

Building on statistical and computational methods, the practice of learning analytics applies a variety of data analysis techniques to describe, characterize, and predict the learning behaviors of individuals.¹ In the clinical realm, emergency physicians are increasingly familiar with data dashboards. Many emergency departments (EDs) have adopted digital interfaces such as patient tracker boards and patient care dashboards.^{2,3} These data provide EDs operational teams with information that is often used to inform logistic workplace-based quality improvement initiatives and guide changes to ensure smooth ED operations and flow.

With the move to competency-based medical education (CBME) via the Accreditation Council for Graduate Medical Education (ACGME) Milestones (2013) and the Royal College of Physician and Surgeons of Canada’s Competence by Design (2017), emergency medicine residency programs have access to ever-increasing amounts of data about the performance of their trainees. However, the systematic application of learning analytics to interpret these data is sporadic. Some groups have started using learning analytics to gather information and gain insights about learner- or system-level

performance (e.g., the McMaster Modular Assessment Program for emergency medicine,^{4–8} online modules that using learning analytics methods to teach x-ray interpretation,^{9,10} and an internal medicine program’s analytics dashboards¹¹). The vast majority of residency programs, however, are attempting to execute programmatic assessment (i.e., the integrated system of multiple, longitudinal observations from multiple observers, aggregated into summary performance scores for group adjudication of global judgment) without optimized data collection (e.g., valid testing/simulations, timely and accurate workplace-based assessments), modern analytic techniques, or appropriate data representation. These tools are essential for the successful implementation and execution of programmatic assessment.^{12–16}

Data are increasingly valued commodities, yet data collection and interpretation increasingly consume faculty time. Efficiencies have been limited by resistance to the amalgamation and security complexities of large data sets, combined with a lack of technical and contextual expertise required for analysis. This challenge has left data sets with considerable potential unanalyzed or underexplored.

THE IMPORTANCE OF THEORY IN LEARNING ANALYTICS

In 2001, Tollock¹⁷ wrote that “If you torture the data long enough, it will confess.” Medical educators would be wise to keep this quote in mind as digitization provides us with ever-increasing amounts of data about our trainees. While larger amounts of data have the potential to inform trainees and their supervisors about learning and progression, they can also facilitate the appearance of meaningless patterns. This occurs both because of “patterns” actually due to chance or particular kinds of measurement error and because interpretations can vary based upon how the data are presented.¹⁸ Furthermore, as people become aware, they may modify their actions or aspects of the system to meet expectations (i.e., gaming the system). Valid interpretations must incorporate educational theory to ensure that useful questions are asked of the data and that the answers are used appropriately.^{19–24}

One of the most important concepts to remember when managing educational data sets is that they are, in fact, a database. That is to say, data are the foundation for the analysis and when handling them, it is important to do so with care and consideration—and only enter into analyses those data points that are theoretically grounded or directed by evidence-based rationales. As

Wise and Shaffer²⁰ question, “What counts as a *meaningful* finding when the number of data points is so large that *something* will always be significant?”

These issues are not new to most clinician educators; when we critically appraise epidemiology database studies, there is an expectation that examined variables are linked in some biologically plausible (theory-driven) or evidence-based (hypothesis-driven) way. Mining data via subgroup analyses is frowned upon in the research community for good reason.²⁵ Since the analytic techniques used in learning analytics are often similar to those used by epidemiologists (e.g., inspired by regression or correlation statistics), unprincipled data mining should likewise be minimized in educational data sets. Figure 2 depicts three places where theory should be integrated within learning analytics.

There is likely a limited role for not-yet-discovered questions which arise from data (which begs for an a posteriori hypothetical approach that big data allows²²); however, such observational analyses must be interpreted with judiciousness. Specifically, recalling the limitations of the techniques used (e.g., remembering that correlations are just measures of association and not causation) will be of utmost importance.

DATA COLLECTION

Educators and clinical supervisors need to be trained to both observe and effectively document their insights on a routine basis. By enhancing the data collection from clinical supervisors, a residency program can enhance their dataset from a dozen data points (e.g., 12 end of rotation reports plus other practice examinations over the course of the year) to several hundred data points (e.g., the McMAP system can gather about 400 data points per resident per year).⁴ System designers and administrators such as program directors will need to consider how they engage in quality assurance and monitoring of their systems. Finally, missing data has very-far-reaching ramifications and should also be considered in these applied systems.⁸

COMMONLY USED LEARNING ANALYTIC TECHNIQUES IN MEDICAL EDUCATION LEARNING

Learning analytics require large sets of machine-readable data.²⁶ Such data sets have been available to other areas of education since the 2000s and before through the collection of data from online learning modules (e.g.,

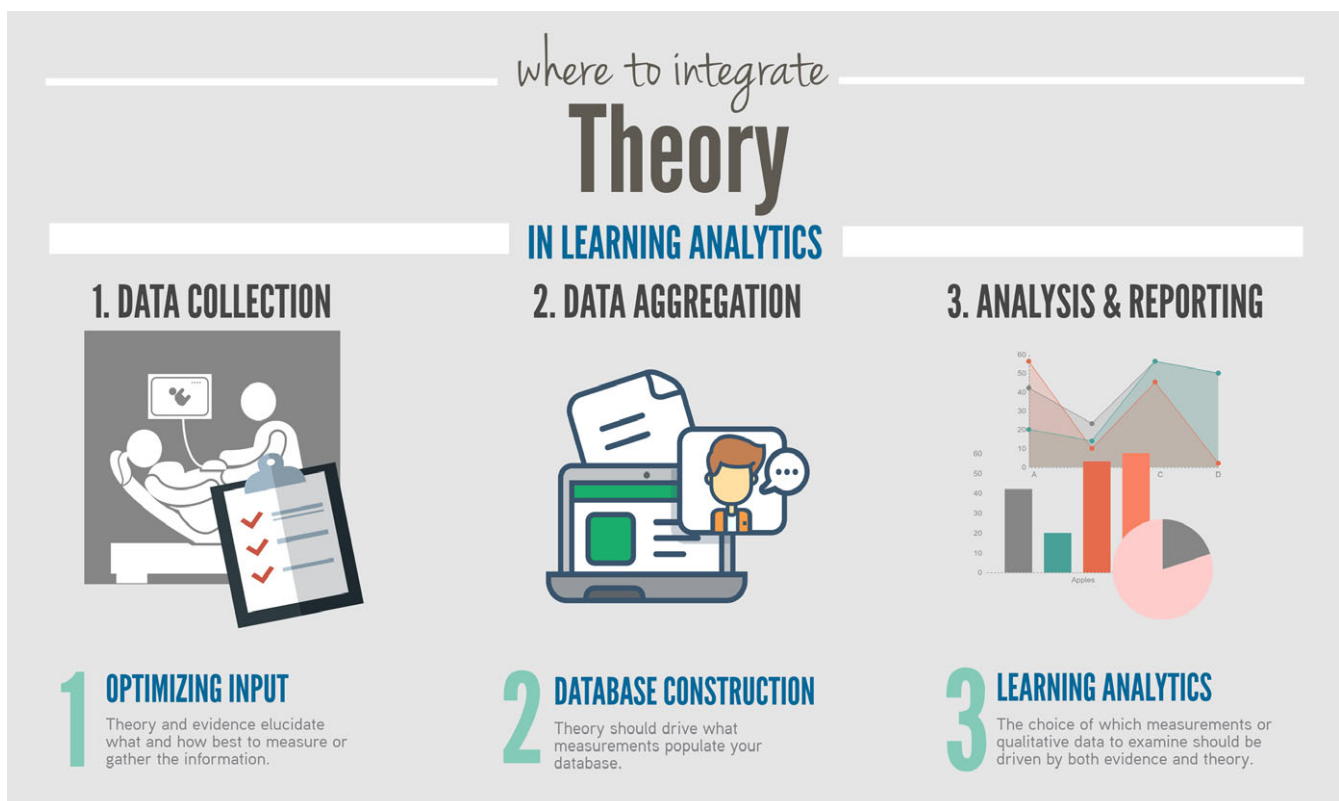


Figure 2. Where to integrate theory in learning analytics.

Table 1
Learning Analytic Categories

Analytic Category	Brief Description	Techniques*	Requirements	Worked or Potential Example in Medical Education
Descriptive analyses of quantitative data	Analyses that characterize past performance or learning process data.	Summary statistics, visual analytics, cluster analysis, social network analysis, heat maps. Currently, this is mainly what medical educators are using to describe their trainees.	Data that can be combined due to their similarity in scaling (i.e., both scores are out of 7) and conceptually (i.e., both items could be aggregated because they are both measuring communication skills).	Time series graphs of average achievement scores (e.g., average in-training exam scores). Radar Plots ³⁹ (graphical representation of multiple data points overlying a grid that marks out various competencies).
Descriptive analyses of qualitative data (i.e., automated text analyses)	Computational analyses of digital resources and/or products created during learning to improve the understanding of learning activities. The automated processing of naturally generated human language.	Computational linguistic analyses of text coherence, syntactic complexity, lexical diversity, semantic similarity. Collaborative filtering techniques, content-based techniques.	Large database of qualitative comments with or without human supervision to allow for primarily text-based processing.	Natural Language Processing (where artificial intelligence is applied to human language) Automated content analysis (where a MLA “learns” to reliably replicate human coding across large volumes of data), e.g., Automated flagging of concerning qualitative comments by faculty based on learned models of textual analysis. [†]
Explanatory modeling	The use of data to identify causal relationships between constructs and outcomes within a data set to refine cognitive models.	Difficulty factors assessment, learning factors analysis, linear or logistic regression, decision trees, automated cognitive model discovery.	Variables that map to defined constructs. Data representative of the target population.	Learning trajectories (regression based), ^{6,11} Learning curves. ⁹ Cumulative sum scores plots. ⁴⁰
Predictive modeling	The use of historical data to generate models that predict unknown future events (e.g., academic success, timeline to program completion) based on timely observations.	Linear and logistic regression, nearest neighbors classifiers, decision trees, naive Bayes classifiers, bayesian networks, support vector machines, neural networks, ensemble methods.	Quantifiable subject characteristics, clear outcome of interest, a large data set, and an ability to intervene. Training and test data for which you know the outcome (e.g., passing boards).	Development of a predictive model for difficulty in residency (e.g., remediation, probation, or failure to complete) for current residents using data from their residency applications. ³⁸ Future unsupervised MLAs may be able to spot patterns in performance without human supervision. [†]
Evaluative analyses	The use of aggregated assessment data can also be flipped on its head and become an analysis of a program’s functioning.	All of the above may be used, the difference is the flip in the item of interest from the trainee toward evaluating the residency program, the faculty, or the system at large.	The ability to link trainee data to the rest of the healthcare or educational system. e.g., if you are evaluating rater gender bias, then your data must be linked to rater data (e.g., the identities of who rated whom when).	A six-program evaluation study of point-of-care daily assessments of milestone data was recently conducted. ^{41,42} This data has opened up a larger conversation about gender bias in EM trainee assessments.

*These techniques are not exclusively used for each type of analysis, but are those which are considered the typical uses. Most techniques can be modified for various uses.

†Starred items represent items which are newer learning analytic techniques on the horizon.

intelligent tutoring system and massive open online course data).²⁶ However, with the digitization of educational records and transition to frequent, direct observation assessments within medical education, databases meeting these criteria are becoming available to medical educators at program, institution, and national levels. In medical education, we rarely have had enough data to use these techniques, but with the increasing use of daily encounter cards,^{27–30} progress testing,³¹ observed clinical encounters,^{32–34} milestones,³⁵ and entrustable professional activity ratings,^{11,36,37} we are afforded opportunities to now employ learning analytics techniques to guide overall assessment decisions. Importantly, learning analytics offer techniques for examining not only the outcomes of learning but the process, providing information that be used to support trainees and improve their learning. Table 1 shows various analytic techniques and how these can help with describing, predicting, explaining, and evaluating assessment data.

ENTER THE MACHINES: MACHINE LEARNING ALGORITHMS AND THEIR ROLE IN LEARNING ANALYTICS

In other realms such as banking or website traffic, data monitoring and pattern recognition are no longer exclusively a human endeavor. Machine learning algorithms (MLAs) are revolutionizing the way that we harness computer technology to recognize, and at times predict, patterns within data. With the advent of MLAs, there are not only inherent opportunities, but also possible dangers. MLAs can find patterns at the broad scale or make inferences about classes of people, but applying these analytics at the individual level is still a challenge. One recent study from Canada showed that MLAs have the potential to assist program directors with screening for residents at risk, but lack the ability to understand trainees' strengths and weaknesses and the sophistication to generate specific remediation plans.³⁸

Nevertheless, MLAs allow large-scale data to inform our understanding of a trainee's progression toward competence and hold great potential as screening tools or early warning systems. For example, real-time data on Zainab's trainees could be continuously analyzed to identify struggling trainees. With adequate data, MLAs have the potential to help educators detect difficulties and problems earlier. More importantly, while learning analytics cannot always ascertain the "why," good data can help you better describe the "what"—i.e., with

good competency data you could describe 1) a trainee's individual strengths and weaknesses or 2) common areas of challenge across a cohort of residents.

AN INTRODUCTION TO LEARNING ANALYTIC TECHNIQUES

Table 1 outlines various analytical categories and associated techniques. Additional techniques that may be of interest to medical educators have also been described in the first edition of the *Handbook of Learning Analytics* by the Society for Learning Analytics Research.¹⁸ Table 2 contains information about how we can apply learning analytics to our medical education assessment data. To complement this table, Data Supplement S1 (available as supporting information in the online version of this paper, which is available at <http://onlinelibrary.wiley.com/doi/10.1002/aet2.10087/full>) contains some examples of learning analytics using an exemplar data set.

WITH GREAT DATA COMES GREAT RESPONSIBILITY: THE CONSEQUENCES AND QUALITY OF DATA

It is important to consider the *consequential validity evidence* of data analyses.⁴³ Described by Messick in 1989,⁴³ the concept of consequential validity has also been explained as educational impact⁴⁴ and refers to the positive or negative consequences of a particular assessment. Just as in other educational scenarios, using a particular construct to make decisions will convey to our trainees that we value that construct.^{43,45} Similarly, when we make decisions about a trainee based on our data analyses, we must also bear in mind the consequences of our decisions affecting their later behaviors and performances, both good and bad. While early identification of trainees at risk may be valuable so programs can offer earlier personalized remediation plans,⁶ the very act of labeling a "trainee-at-risk" may have consequences toward their self-perceptions and future performance.^{46,47} At the same time, the implications of *not* collecting and interpreting data may also be problematic. Without timely data inputs (e.g., continuous workplace-based assessment data from every day/shift in a rotation), we may be glossing over problems via acts of *omission*. It is critical, then, that those who are charged with aggregating and interpreting larger swaths of data to be aware of potential pitfalls of missing data⁸ or cognitive biases

Table 2
How This Applies to Your Learning Environment

Role	How Learning Analytics Might Affect Educators and Clinical Supervisors in Their Various Roles	Examples of Applying Learning Analytics in Medical Education (Based on the Initial Case Vignette)
<p>Clinical supervisors</p> <p>At the bedside, teachers will often be the main source of observations. Historically, they have been the main source of data about clinical performance, but usually they are asked to supply anecdotal evidence, retrospectively (i.e., in-training or month-end reports). The learning analytics age will likely usher in new requirements and expectations for this group. In systems embracing the analytics age, they are often asked via daily encounter cards to jot down scores and short notes about resident performance, which can then be analyzed in aggregate or to identify trajectories over time.</p>	<p>Dr. Joe Santiago (a first-year faculty member) is often overwhelmed by clinical care and supervisory responsibilities and quite frankly often forgets to complete the mandated daily assessments for the trainees that pass through. He is working on being a better citizen of the system, reserving some time at the beginning, middle, and end of shift to perform direct observations, take down feedback, and then transcribe that feedback into their online trainee assessment system. However, recently he also attended a faculty development after reading local quality assurance data about how female residents receive qualitatively different feedback than male counterparts. He considers using more specific language that focuses on observed actions for both male and female trainees—steering away from engendered stereotypes and vague generalizations.</p>	<p>Dr. Sarah Patterson is the PD and has been working with the software company that runs their institutional residency assessment system to optimize the emergency medicine daily encounter cards. After reading the recent series of papers about gender bias in locally derived milestone data, she is also working with the software company to find a way to extract faculty level data from the local database to examine if there are gender biases within their local context. She has also created a task force that is creating a series of faculty development workshops to enhance the quality of comments written by faculty members in the daily encounter card system.</p>
<p>Systems administrators, managers, and designers, e.g., PDs, APDs</p> <p>Those who are most responsible for the data administration and management of trainee assessment data often are part of the program executive. PDs and APDs will often serve as both designers and managers of these data systems and should remain vigilant of systemic problems that might best be uncovered through routine scrutiny of the data generated by faculty members. They are in charge of quality assurance for both the data and the system, ensuring that the faculty development required to train raters in the system are sufficient, but also may need to run analytics to generate program evaluation data to see if there are systematic biases.</p>	<p>Dr. Zainab Hussain decides to create a new template system to input the data so that she can present her residents more effectively to their local competency committee. She writes an e-mail to the program APD when she reviews the files and notes that there may be a systematic problem with the way that international medical graduates are assessed within their system. Now that they have access to nearly 200 data points per resident per year in their local system, she also suggests to the program director that they look into some MLAs (e.g., natural language processing and predictive modeling) to make better use of the data that they currently gather. Together, she and the PD submit a grant application to a local innovation fund, seeking money to purchase and pilot this innovation.</p>	<p>Dr. Wei-Ling Yee is assigned as a direct observer of a PGY1 trainee who has been flagged by the competency committee as a resident who is struggling with knowledge synthesis. After receiving handover from the PD (Dr. Patterson), and permission from the trainee, he was given full access to the trainee's file. Armed with the trainees first 3 months of learning data, Dr. Yee has also spent time working with the at-risk resident, following her around two shifts now with a 1:1 educator: trainee ratio. He has noted a pattern in the trainee's ability to gather data during the initial history and physical encounter. As a team, they have set a goal to improve this, and he will be recommending a series of simulation sessions with standardized patients to focus on this data collection skill. They both intend to continue monitoring her learning trajectory over the next 6 months, meeting regularly to see if there are improvements. As a more senior member of the faculty, he also notes that several midcareer faculty members are writing noncontributory comments (e.g., "read more," "great job") and has offered to help out with future faculty development workshops on improving this skill set.</p>
<p>Meta-raters CCC portfolio reviewer</p> <p>At this point, meta-raters such as competency committee portfolio reviewers are a crucial part of the system. These individuals review all the data available to them, then synthesize and present these data to the competency committees for decision. Learning analytics have the chance to greatly decrease the burden on these individuals, decreasing the workload that they will have. Although it would seem that learning analytics might make these individuals obsolete, learning analytics as manifest via machine learning or artificial intelligence will largely need to be built on the cognitive models of these individuals. But models are never perfect, just the best approximation we can currently build. Thus, even with automatic processing of human language and rater-derived data, human interpretation and oversight will always be needed.</p>	<p>Coaches and those responsible for remediation will have the opportunity to review data that they have never had before. Some programs will have formal advisor programs, but for other programs these individuals will likely emerge as informal mentorship roles. These increased data hold great potential to naturally enhance the work that these individuals do—which is to seek out the weaknesses of a trainee and help them to bolster their knowledge or skills in those areas. However, as trainees have access to more and more individualized data, it will be increasingly important for this group of individuals to help guide learners through interpreting the data and then helping them to set reasonable and proximate goals for improvement. For advanced trainees, this may mean that their coach will help them to find new challenges. For trainees with focal difficulties, this may require targeted planning around properly pinpointing the difficulty and enacting a personalized learning plan based on a trainee's data.</p>	<p>Dr. Wei-Ling Yee is assigned as a direct observer of a PGY1 trainee who has been flagged by the competency committee as a resident who is struggling with knowledge synthesis. After receiving handover from the PD (Dr. Patterson), and permission from the trainee, he was given full access to the trainee's file. Armed with the trainees first 3 months of learning data, Dr. Yee has also spent time working with the at-risk resident, following her around two shifts now with a 1:1 educator: trainee ratio. He has noted a pattern in the trainee's ability to gather data during the initial history and physical encounter. As a team, they have set a goal to improve this, and he will be recommending a series of simulation sessions with standardized patients to focus on this data collection skill. They both intend to continue monitoring her learning trajectory over the next 6 months, meeting regularly to see if there are improvements. As a more senior member of the faculty, he also notes that several midcareer faculty members are writing noncontributory comments (e.g., "read more," "great job") and has offered to help out with future faculty development workshops on improving this skill set.</p>

APD = assistant program directors; CCC = clinical competency committee; PD = program director.

(e.g., halo effect⁴⁸) that may affect the meta-rating⁶ (i.e., interpretation of multiple ratings from multiple raters) of the presented data.

Indeed, it is key to monitor our data systems for possible biases. While applying data-driven, actuarial approaches as described in Table 1 can assist individual meta-raters to better interpret individual trainee performance,⁴⁸ spinning the data set on its head to place the raters under the microscope allows us to use evaluative analyses to examine our assessment systems. Recent literature has shown that, based on the present EM milestones, there may be a significant difference in the way that female trainees are rated when compared to their male counterparts.^{41,42} Studies like these recent ones in our own specialty,^{41,42} as well as prior educational measurement literature,^{49–51} remind us of how important it is to check assessment systems for sources of implicit and hidden bias. The aggregation of data in these studies has allowed for powerful analyses that have brought these thought-provoking issue to light and may also illuminate other systemic issues.^{41,42} If a problem occurs at one site, you may wonder about a specific group's local culture, but if a problem occurs across multiple programs, then systemic biases may be at play and may warrant reexamination of the system for flaws. Is the data acquisition processes

flawed? Is there insufficient faculty development for rater training, analysis of terminology, and language used in assessment systems? One exciting value of learning analytics for educators is the ability to monitor their data in a post hoc manner (i.e., the concept of a *a posteriori* hypotheses²²) to detect concerning patterns which can prompt further investigations and improvements.

THE FUTURE

Increasing the yield of educationally informative analyses from CBME is critical. The significant cost and resources required to transform our residency training systems requires a return on that investment. Targeted analyses of trainees and systems demonstrate increased accountability being asked of medical education by our leaders, our patients, and our trainees.^{52,53}

Larger-scale data sharing may also allow us to complete analyses that can inform broader policy decisions (e.g., human resources planning, funding).⁶ For instance, if we can predict how long it takes for a typical trainee to progress through a system compared to how a gifted or struggling trainee might progress, we can begin to anticipate what sort of additional funding or new training opportunities are needed to ensure

Table 3
Key References About Learning Analytics for Clinician Educators

Citation	Why This Paper Is Important
Key conceptual papers	
Bok HG, Teunissen PW, Favier RP, et al. Programmatic assessment of competency-based workplace learning: when theory meets practice. <i>BMC Med Educ</i> 2013;13:123. ¹³	This paper provides a good overview of what the concept programmatic assessment means and how it can help to lay basis for mapping out the type of assessment data that on might wish to gather.
Pusic MV, Boutis K, Hatala R, Cook DA. Learning curves in health professions education. <i>Acad Med</i> 2015;90:1. ⁹	This provides a conceptual overview about learning curves and how they might apply to health professions education.
Cirigliano MM, Guthrie C, Pusic MV, et al. "Yes, and . . ." Exploring the future of learning analytics in medical education. <i>Teach Learn Med</i> 2017;29:368–72. ¹²	This is a short paper that highlights the opportunities for using learning analytics in medical education.
Key implementation-related papers	
Chan TM, Sherbino J, Mercuri M. Nuance and noise: lessons learned from longitudinal aggregated assessment data. <i>J Grad Med Educ</i> 2017;9:724–9. ⁶	This is an example of how assessment data might be aggregated and mapped, explaining to educators how this data might be harnessed, and what noise might surround the emergent data.
Ginsburg S, Regehr G, Lingard L, Eva KW. Reading between the lines: faculty interpretations of narrative evaluation comments. <i>Med Educ</i> 2015;49:296–306. ⁵⁴	This paper highlights the importance of qualitative data and how administrators and competency committee members might interpret these data.
McConnell M, Sherbino J, Chan TM. Mind the gap: the prospects of missing data. <i>J Grad Med Educ</i> 2016;8:708–12. ⁸	This paper highlights a systems-level analysis of how administrators and competency committee members might consider "missing data." It provides a suggestion of applying a framework similar to Yvonne Steinert's KSALTS framework to analyze the problem that missing data presents.
Warm EJ, Held JD, Hellmann M, et al. Entrusting observable practice activities and milestones over the 36 months of an internal medicine residency. <i>Acad Med</i> 2016;91:1398–405. ¹¹	This is a worked example of a residency program with a large number of residents and how they visualized and used the data they acquired about workplace-based assessments for educational purposes.

adequate training for all. Identifying a gifted resident may allow a program director to begin negotiating a critical care rotation earlier for that talented young physician. Meanwhile, a struggling trainee may need extra time in a simulation lab or repetition of a clinical experience for more practice (e.g., an extra anesthesia block for focused airway management skills). Personalized education paths can better meet individual needs, and we also must caution against systems so they do not overly constrain trainees' potential trajectories prematurely. "Labeling" trainees based solely on their early performance trends may prove to be more detrimental than advantageous.

One of the most important untapped features of many of our current assessment systems will be the analysis of the qualitative (i.e., narrative feedback) data that are generated by faculty members. Analysis of rich qualitative data promises a better understanding of context and the nuance of performance that arise in the clinical sphere. Recent work has shown that written comments can be used to reliably discriminate between high and low performing trainees.^{54,55} Text-based applications of MLAs (e.g., natural language processing, automated content analyses) or other keyword-based analytic tools may be useful in helping us to better visualize, interpret, and leverage the rich information contained within qualitative data. Moreover, the combined interpretation of both qualitative and quantitative traces of trainees' assessment may help to identify gaps in the global analysis.

HOW CAN I LEARN MORE?

Learning and assessment analytics is an exciting area of health professions education that has great potential to shape the way we deliver, improve, and customize our trainee experiences. However, getting started in this field is admittedly difficult. To help you get started, we have compiled some resources (Table 3) that can help educators who are interested in reading more around learning analytics; many more references are contained in the references to this paper. For those interested in learning about how to process and aggregate their local data, it is useful to consider attending international medical education conferences (e.g., International Conference on Residency Education, Ottawa Conference on Assessment, the International Competency Based Medical Education preconference summit before the Association for Medical Education in Europe) where there are more

frequently workshops about data-driven assessments. For those who are looking for some of these newer learning analytics techniques, advanced training through courses on machine learning via massive open online courses or the Society of Learning Analytics Research (<https://solarrsearch.org>) may also be useful.

CASE RESOLUTION

Reviewing the learning analytics literature, Zainab decides to create a radar graph of each of her trainees' strengths, weaknesses, and areas for further development. With the help of a colleague, she applies a natural language processing algorithm to sort through the large volume of qualitative data reports, highlighting the few comments of concern or confusion out of the hundreds submitted. She shares her templates with the rest of the competency committee members. Knowing that some of these techniques are new to the discipline of emergency medicine education, Zainab speaks with her program director and local research scientist about setting up a new system to continuously survey and improve the process and produce scholarship along the way to help others find best practices.

We thank Dr. Susan Promes and the AEM Education and Training editorial board for inspiring this piece. Drs. Chan and Sherbino thank Drs. Alim Pardhan, Mathew Mercuri, and Ian Preyra for continuing to support the McMAP program and all of their local colleagues at McMaster University for their tireless pursuit improving resident performance. Dr. Pusic thanks his collaborators Drs. Kathy Boutis and Martin Pecaric for their work in the ImageSim program.

References

1. Cooper A. What is analytics? Definition and essential characteristics. *CETIS Anal Ser* 2012;1:1–10.
2. Aronsky D, Jones I, Lanaghan K, Slovis CM. Supporting patient care in the emergency department with a computerized whiteboard system. *J Am Med Inform Assoc* 2008;15:184–94.
3. Dowding D, Randell R, Gardner P, et al. Dashboards for improving patient care: review of the literature. *Int J Med Inform* 2015;84:87–100.
4. Chan T, Sherbino J; McMAP Collaborators. The McMaster Modular Assessment Program (McMAP): a theoretically grounded work-based assessment system for an emergency medicine residency program. *Acad Med* 2015;90:900–5.
5. Li S, Sherbino J, Chan TM. McMaster Modular Assessment Program (McMAP) through the years: residents'

- experience with an evolving feedback culture over a 3-year period. *AEM Educ Train* 2017;1:5–14.
6. Chan TM, Sherbino J, Mercuri M. Nuance and noise: lessons learned from longitudinal aggregated assessment data. *J Grad Med Educ* 2017;9:724–9.
 7. Sebok-Syer SS, Klinger DA, Sherbino J, Chan TM. Mixed messages or miscommunication? investigating the relationship between assessors? Workplace-based assessment scores and written comments. *Acad Med* 2017;92:1774–9.
 8. McConnell M, Sherbino J, Chan TM. Mind the gap: the prospects of missing data. *J Grad Med Educ* 2016;8:708–12.
 9. Pusic MV, Boutis K, Hatala R, Cook DA. Learning curves in health professions education. *Acad Med* 2015;90:1.
 10. Pusic M, Pecaric M, Boutis K. How much practice is enough? Using learning curves to assess the deliberate practice of radiograph interpretation. *Acad Med* 2011;86:731–6.
 11. Warm EJ, Held JD, Hellmann M, et al. Entrusting observable practice activities and milestones over the 36 months of an internal medicine residency. *Acad Med* 2016;91:1398–405.
 12. Schuwirth LW, van der Vleuten CP. Programmatic assessment and Kane's validity perspective. *Med Educ* 2012;46:38–48.
 13. Bok HG, Teunissen PW, Favier RP, et al. Programmatic assessment of competency-based workplace learning: when theory meets practice. *BMC Med Educ* 2013;13:123.
 14. Schuwirth LW, Van der Vleuten CP. Programmatic assessment: from assessment of learning to assessment for learning. *Med Teach* 2011;33:478–85.
 15. van der Vleuten CP, Schuwirth LW, Driessen EW, et al. A model for programmatic assessment fit for purpose. *Med Teach* 2012;34:205–14.
 16. Van Der Vleuten CP, Schuwirth LW, Driessen EW, Govaerts MJ, Heeneman S. Twelve tips for programmatic assessment. *Med Teach* 2015;37:641–6.
 17. Tollock G. A comment on Daniel Klein's "A Plea to Economists Who Favor Liberty". *East Econ J* 2001;27:203–7.
 18. Lang C, Siemens G, Wise A, Gašević D, editors. *Handbook of Learning Analytics*. 1st ed. Ann Arbor (MI): Society for Learning Analytics Research, 2017.
 19. Knight S, Buckingham Shum S, Littleton K. Epistemology, assessment, pedagogy: where learning meets analytics in the middle space. *J Learn Anal* 2014;1:23–47.
 20. Wise AF, Shaffer DW. Why theory matters more than ever in the age of big data. *J Learn Anal* 2015;2:5–13.
 21. Cirigliano MM, Guthrie C, Pusic MV, et al. "Yes, and ..." Exploring the future of learning analytics in medical education. *Teach Learn Med* 2017;29:368–72.
 22. Ellaway RH, Pusic MV, Galbraith RM, Cameron T. Developing the role of big data and analytics in health professional education. *Med Teach* 2014;36:216–22.
 23. Ellaway R, Pusic M, Galbraith R, Cameron T. Re: "Better data >> bigger data". *Med Teach* 2014;36:1009.
 24. Pusic MV, Boutis K, McGaghi WC. Role of scientific theory in simulation education research. *Simul Healthc* 2018 [Epub ahead of print].
 25. Guyatt GH, Oxman AD, Kunz R, et al. GRADE guidelines: 7. Rating the quality of evidence - inconsistency. *J Clin Epidemiol* 2011;64:1294–302.
 26. Ferguson R. Learning analytics: drivers, developments and challenges. *Int J Technol Enhanc Learn* 2012;4:304.
 27. Bandiera G, Lendrum D. Daily encounter cards facilitate competency-based feedback while leniency bias persists. *Can J Emerg Med*. 2008;10:44–50.
 28. Sherbino J, Kulasegaram K, Worster A, Norman GR. The reliability of encounter cards to assess the CanMEDS roles. *Adv Heal Sci Educ* 2013;18:987–96.
 29. Pelgrim EA, Kramer AW, Mokkink HG, van den Elsen L, Grol RP, van der Vleuten CP. In-training assessment using direct observation of single-patient encounters: a literature review. *Adv Heal Sci Educ* 2011;16:131–42.
 30. Sokol-Hessner L, Shea JA, Kogan JR. The open-ended comment space for action plans on core clerkship students' encounter cards: what gets written? *Acad Med* 2010;85:S110–4.
 31. Schiff K, Williams DJ, Pardhan A, Preyra I, Li SA, Chan T. Resident development via progress testing and test-marking: an innovation and program evaluation. *Cureus* 2017;9:1–8.
 32. Moonen-van Loon JM, Overeem K, Donkers HH, van der Vleuten CP, Driessen EW. Composite reliability of a workplace-based assessment toolbox for postgraduate medical education. *Adv Heal Sci Educ* 2013;18:1087–102.
 33. Norcini JJ, Blank LL, Duffy FD, Fortna GS. Academia and clinic the mini-CEX: a method for assessing clinical skills. *Ann Intern Med* 2003;138:476–81.
 34. Hayward M, Chan T, Healey A. Dedicated time for deliberate practice: one emergency medicine program's approach to point-of-care ultrasound (PoCUS) training. *CJEM* 2015;17:558–61.
 35. Korte RC, Beeson MS, Russ CM, Carter WA, Reisdorff EJ. The emergency medicine milestones: a validation study. *Acad Emerg Med* 2013;20:730–5.
 36. van Loon KA, Driessen EW, Teunissen PW, Scheele F. Experiences with EPAs, potential benefits and pitfalls. *Med Teach* 2014;36:698–702.
 37. Sterkenburg A, Barach P, Kalkman C, Gielen M, ten Cate O. When do supervising physicians decide to entrust residents with unsupervised tasks? *Acad Med* 2010;85:1408–17.
 38. Ariaeinejad A, Samavi R, Chan T, Doyle T. A performance predictive model for emergency medicine residents. In: *27th Annual International Conference on Computer Science and Software Engineering*. Toronto, ON: ACM Digital Library, 2017.

39. Warm EJ, Schauer D, Revis B, Boex JR. Multisource feedback in the ambulatory setting. *J Grad Med Educ* 2010;2:269–77.
40. Peyrony O, Legay L, Morra I, et al. Monitoring personalized learning curves for emergency ultrasound with risk-adjusted learning-curve cumulative summation method. *AEM Educ Train* 2018;2:10–4.
41. Dayal A, O'Connor DM, Qadri U, Arora VM. Comparison of male vs female resident milestone evaluations by faculty during emergency medicine residency training. *JAMA Intern Med* 2017;177:651.
42. Mueller AS, Jenkins T, Osborne M, Dayal A, O'Connor DM, Arora VM. Gender differences in attending physicians' feedback for residents in an emergency medical residency program: a qualitative analysis. *J Grad Med Educ* 2017;9:577–85.
43. Messick S. Validity. In: Linn RL, editor. *Educational Measurement*. 3rd ed. Austin (TX): Macmillan, 1989.
44. Hamdy H. AMEE Guide Supplements: workplace-based assessment and an educational tool. Guide supplement 31.1–viewpoint. *Med Teach* 2009;31:59–60.
45. Van Der Vleuten CP, Schuwirth LW, Scheele F, Driessen EW, Hodges B. The assessment of professional competence: building blocks for theory development. *Best Pract Res Clin Obstet Gynaecol* 2010;24:703–19.
46. Cleland J, Arnold R, Chesser A. Failing finals is often a surprise for the student but not the teacher: identifying difficulties and supporting students with academic difficulties. *Med Teach* 2005;27:504–8.
47. Holland C. Critical review: medical students' motivation after failure. *Adv Heal Sci Educ* 2016;21:695–710.
48. Sherbino J, Norman G. On rating angels: the halo effect and straight line scoring. *J Grad Med Educ* 2017;9:721–3.
49. Wind SA, Sebok-Syer S. Infit, outfit, and between fit statistics for raters: The difference is in the details. Paper presented at the annual meeting of the Canadian Society for the Study of Education, Toronto, Ontario, Canada, May 2017.
50. Smith R. Thoughts for new medical students at a new medical school. *BMJ* 2003;327:1430–3.
51. Ackerman TA. A didactic explanation of item bias, item impact, and item validity from a multidimensional perspective. *J Educ Meas* 1992;29:67–91.
52. Taber S, Frank JR, Harris KA, Glasgow NJ, Iobst W, Talbot M. Identifying the policy implications of competency-based education. *Med Teach* 2010;32:687–91.
53. Caccia N, Nakajima A, Kent N. Competency-based medical education: the wave of the future. *J Obstet Gynaecol Canada* 2015;37:349–53.
54. Ginsburg S, Regehr G, Lingard L, Eva KW. Reading between the lines: faculty interpretations of narrative evaluation comments. *Med Educ* 2015;49:296–306.
55. Ginsburg S, van der Vleuten CP, Eva KW, Lingard L. Cracking the code: residents' interpretations of written assessment comments. *Med Educ* 2017;51:401–10.

Supporting Information

The following supporting information is available in the online version of this paper available at <http://onlinelibrary.wiley.com/doi/10.1002/aet2.10087/full>

Data Supplement S1. Examples of Learning Analytics.