Canadian Medical Education Journal
Revue canadienne de l'éducation médicale

Developing a dashboard for program evaluation in competency-based training programs: A design-based research project
Création d'un tableau de bord pour l'évaluation des programmes de formation fondée sur les compétences : un projet de recherche basé sur la conception

Yusuf Yilmaz, Robert Carey, Teresa M Chan, Venkat Bandi, Shisong Wang, Robert A Woods, Debajyoti Mondal and Brent Thoma

Cite this article

Article abstract

Background: Canadian specialist residency training programs are implementing a form of competency-based medical education (CBME) that requires the assessment of entrustable professional activities (EPAs). Dashboards could be used to track the completion of EPAs to support program evaluation.

Methods: Using a design-based research process, we identified program evaluation needs related to CBME assessments and designed a dashboard containing elements (data, analytics, and visualizations) meeting these needs. We interviewed leaders from the emergency medicine program and postgraduate medical education office at the University of Saskatchewan. Two investigators thematically analyzed interview transcripts to identify program evaluation needs that were audited by two additional investigators. Identified needs were described using quotes, analytics, and visualizations.

Results: Between July 1, 2019 and April 6, 2021 we conducted 17 interviews with six participants (two program leaders and four institutional leaders). Four needs emerged as themes: tracking changes in overall assessment metrics, comparing metrics to the assessment plan, evaluating rotation performance, and engagement with the assessment metrics. We addressed these needs by presenting analytics and visualizations within a dashboard.

Conclusions: We identified program evaluation needs related to EPA assessments and designed dashboard elements to meet them. This work will inform the development of other CBME assessment dashboards designed to support program evaluation.

Developing a dashboard for program evaluation in competency-based training programs: a design-based research project

Création d'un tableau de bord pour l'évaluation des programmes de formation fondée sur les compétences : un projet de recherche basé sur la conception

Yusuf Yilmaz,[1,2] Robert Carey,[3] Teresa M Chan,[1,4] Venkat Bandi,[5] Shisong Wang,[5] Robert A Woods,[3] Debajyoti Mondal,[5] Brent Thoma[3,6]

[1]Continuing Professional Development Office, and McMaster program for Education Research, Innovation, and Theory (MERIT), McMaster University, Ontario, Canada; [2]Department of Medical Education, Ege University, Turkey; [3]Department of Emergency Medicine, University of Saskatchewan, Saskatchewan, Canada; [4]Division of Emergency Medicine, Department of Medicine at McMaster University; [5]Department of Computer Science, University of Saskatchewan, Saskatchewan, Canada; [6]Royal College of Physicians and Surgeons of Canada, Ontario, Canada
Correspondence to: Dr. Brent Thoma, Room 2646, Box 16, 103 Hospital Drive, Saskatoon, SK. S7N 0W8; email: brent.thoma@usask.ca; Cell: 306 881 0112; Twitter: @Brent_Thoma

Abstract

Background: Canadian specialist residency training programs are implementing a form of competency-based medical education (CBME) that requires the assessment of entrustable professional activities (EPAs). Dashboards could be used to track the completion of EPAs to support program evaluation.

Methods: Using a design-based research process, we identified program evaluation needs related to CBME assessments and designed a dashboard containing elements (data, analytics, and visualizations) meeting these needs. We interviewed leaders from the emergency medicine program and postgraduate medical education office at the University of Saskatchewan. Two investigators thematically analyzed interview transcripts to identify program evaluation needs that were audited by two additional investigators. Identified needs were described using quotes, analytics, and visualizations.

Results: Between July 1, 2019 and April 6, 2021 we conducted 17 interviews with six participants (two program leaders and four institutional leaders). Four needs emerged as themes: tracking changes in overall assessment metrics, comparing metrics to the assessment plan, evaluating rotation performance, and engagement with the assessment metrics. We addressed these needs by presenting analytics and visualizations within a dashboard.

Conclusions: We identified program evaluation needs related to EPA assessments and designed dashboard elements to meet them. This work will inform the development of other CBME assessment dashboards designed to support program evaluation.

Résumé

Contexte : Les programmes de résidence dans une spécialité au Canada offrent une formation médicale fondée sur les compétences (FMFC) qui exige l'évaluation des activités professionnelles confiables (APC). Des tableaux de bord pourraient être utilisés pour suivre la complétion des APC afin de faciliter l'évaluation des programmes.

Méthodes : Par un processus de recherche basé sur la conception, nous avons ciblé les besoins des programmes liés aux évaluations de la FMFC et conçu un tableau de bord qui comprend les éléments (données, analyses et visualisations) nécessaires pour répondre à ces besoins. Nous avons eu recours à des entretiens auprès des responsables du programme de médecine d'urgence et du bureau d'éducation médicale postdoctorale de l'Université de Saskatchewan. Deux enquêteurs ont effectué une analyse thématique des transcriptions des entretiens afin de recenser les besoins d'évaluation du programme, vérifiés par la suite par deux autres enquêteurs. Les besoins recensés ont été décrits à l'aide de citations, d'analyses et de visualisations.

Résultats : Entre le 1er juillet 2019 et le 6 avril 2021, nous avons mené 17 entretiens avec six participants (deux responsables de programmes et quatre responsables de l'établissement). Quatre besoins sont ressortis en tant que thèmes : le suivi des changements dans les mesures d'évaluation globales, la comparaison des mesures avec le plan d'évaluation, l'évaluation de l'efficacité du stage et l'engagement face à des mesures d'évaluation. Nous avons répondu à ces besoins en présentant des analyses et des visualisations dans un tableau de bord.

Conclusions : Nous avons identifié les besoins d'évaluation du programme liés aux évaluations des APC et conçu des éléments de tableau de bord pour y répondre. Ce travail guidera la conception d'autres tableaux de bord d'évaluation de la FMFC en vue de faciliter l'évaluation des programmes.

Introduction

The Royal College of Physicians and Surgeons of Canada began implementing Competency Based Medical Education (CBME) with a programmatic assessment model[1,2] called *Competence By Design* (CBD) in 2016.[3] Within this new paradigm of postgraduate training, faculty complete frequent, low-stakes assessments of entrustable professional activities (EPAs; essential tasks of a discipline that a learner can be trusted to perform[4]) while observing residents' work.[5] These observations include an entrustment score[6–8] and narrative feedback which are organized across four stages of training (transition to discipline, foundations of discipline, core of discipline, and transition to practice) of variable length. The national specialty committee for each discipline both developed the EPAs and provided guidance on the number of each EPA that must be observed.[9] As part of the implementation of CBD, programs created curriculum maps which assign each EPA to appropriate rotations or educational experiences.[9,10] For instance, the specialty of emergency medicine (EM) requires 510 EPA assessments across 28 EPAs for each trainee.[9,11]

The evaluation of competency-based assessment programs is challenging due to the variability in each program's context and implementation.[12] For example, programs situated in community versus academic hospital environments may have different exposures to selected EPAs. Additionally, the shift from a process-oriented to outcomes-oriented educational program has challenged educators as they struggle to identify and measure appropriate educational and clinical outcomes.[13–16] There was limited data to inform an appropriate target number of each EPA assessment prior to gaining competence.[9] The substantially increased volume of assessments[17] provide an opportunity to investigate some educational outcomes. However, analyzing and visualizing this data to support resident assessment is challenging[18–20] and compiling the assessment data from multiple learners to support program evaluation is even more difficult.[21]

Analytical and visualization techniques have been developed in business and sports that support data-driven decision making[22–24] and can be presented in interactive dashboards.[25,26] The use of these techniques with program-level EPA data could support program evaluation by providing insight into the implementation of the prescribed CBD assessment system. At present, there are no evidence-informed descriptions of the analysis or visualization of CBME assessment data for program evaluation.

Our previous work used a design-based research process[27–29] to identify resident,[19] competence committee,[20] and faculty development[30] needs and created dashboards to meet them. We used a similar process to identify and address program evaluation needs.

Methods

Study purpose

The purpose of this study was to identify program evaluation needs for the analysis and presentation of competency-based assessments while developing a dashboard to meet them. We employed an iterative, design-based research process[27–29] and followed best practices in dashboard design.[31,32] Our research methodology was deemed exempt from ethical review by the University of Saskatchewan Research Ethics Board (BEH ID 1655).

Setting

This project was situated within the Department of EM and the office of postgraduate medical education at the University of Saskatchewan between July 1, 2019 and June 30, 2021. The Royal College EM residency was our program of interest. It transitioned to the EPA-based *Competence By Design* program in July of 2018.[9] By academic year, our program had 18 residents enrolled in 2019-20 and 17 residents enrolled in 2020-21.

Participants

We conducted 16 interviews with six participants between July 9, 2019 and April 6, 2021 that ranged in length from 20 to 51 minutes (average of 34 minutes). While this is a relatively small number of participants for a qualitative study, this is in keeping with the research methodology which requires researchers to focus on the individuals involved in the use of the tool or experience that is being developed.[27,33] Participants were selected because of their role in directly conducting and/or overseeing the implementation and evaluation of the assessment program. They included two EM residency program leaders (four interviews with the Program Director and three with the Associate Program Director) and nine interviews with four institutional leaders (four with the Associate Dean of Postgraduate Medical Education, one with the Vice Dean of Education, three with the primary institutional CBD lead, and one with a second institutional CBD lead). All participants were asked via email by the senior author (BT)

to participate in interviews in-person or via video conference. The interviews were conducted by the senior author (BT).

Study protocol

The design-based research methodology[27,29,33] that we employed aligned with the approach outlined in our previous work on competency committee, resident, and faculty dashboards.[19,20,30] Design-based research is an "authentic, contextually aware, collaborative, theoretically focused, methodologically diverse, practical, iterative, and operation-oriented" process[27] which aims to bridge practice and research in education through the integration of investigation and intervention.[27,29,33] In this case, the research component of our study aimed to thematically identify program evaluation needs while addressing them within a dashboard. Figure 1 contains an overview of the four phases of design-based research as they were conducted within our process.[27,29,34]

Phase 1. Analysis and exploration

The senior author (BT) reviewed the literature on program evaluation in competency-based medical education,[13–15,17,18,35,36] learning analytics,[21,37–39] and data visualization[25,31,37,40] to generate ideas for the initial iteration of the program evaluation dashboard. The initial interviews occurred with the program director, associate program director, institutional CBD Lead, postgraduate Dean. They were asked:

1) To outline the information that they use to evaluate their competency-based assessment program.

2) To describe the essential elements of a dashboard containing the program evaluation information they needed.

The participants' responses were further investigated with probing and clarifying follow-up questions.

Phase 2. Design and construction

The initial interview data were transcribed and qualitatively analyzed to inform the creation of a dashboard prototype. The themes from these sessions were reviewed and described to the programmers (VB and SW) who designed a prototype dashboard (Phase 2.1 in Figure 1). The dashboard prototype was released to the program leaders in October of 2019. As described in Phase 3, preliminary analyses of further interviews where participants evaluated and reflected on their use of the prototype were fed back into the design and construction process (Phase 2.2 in Figure 1).

Phase 3. Evaluation and reflection

Phases 2.2 and 3 continued to alternate into the next year. Follow up interviews were held with each of the program and institutional leaders that were interviewed in Phase 1 as well as additional institutional leaders including a second institutional CBD Lead and the Vice-Dean of Education. Interviews occurred more frequently with the program director and associate program director as they had more direct feedback regarding the dashboard. Interviews were scheduled after participants had been given access to the dashboard. During the interview, participants were asked to comment on each of the dashboard elements as they scrolled through them. They were specifically prompted to comment on whether information they required was missing, difficult to access, or difficult to understand within both the individual elements and the entire dashboard. Following each of the interviews, the narrative data was transcribed and qualitatively analyzed to inform the development of the thematic framework and the evolution of the dashboard (Phase 2.2 in Figure 1).

Phase 4. Implementation and spread

The final phase of Design-Based Research describes the implementation and spread of the innovation. Notably, the dashboard is published online under an open access license[41] and has been used by 13 other programs at the University of Saskatchewan (12 other postgraduate programs and the undergraduate program). It is also being adapted for use within an international Learning Management System (Elentra, Elentra Consortium, Kingston, ON). While this demonstrates the impact of our dashboard beyond the context of this work within our EM training program, this phase did not contribute directly to the determination of program evaluation needs, so we do not describe it in our results. We anticipate that this work will continue as the dashboard is implemented more broadly, with implementation challenges being formally or informally evaluated and addressed within the dashboard design.
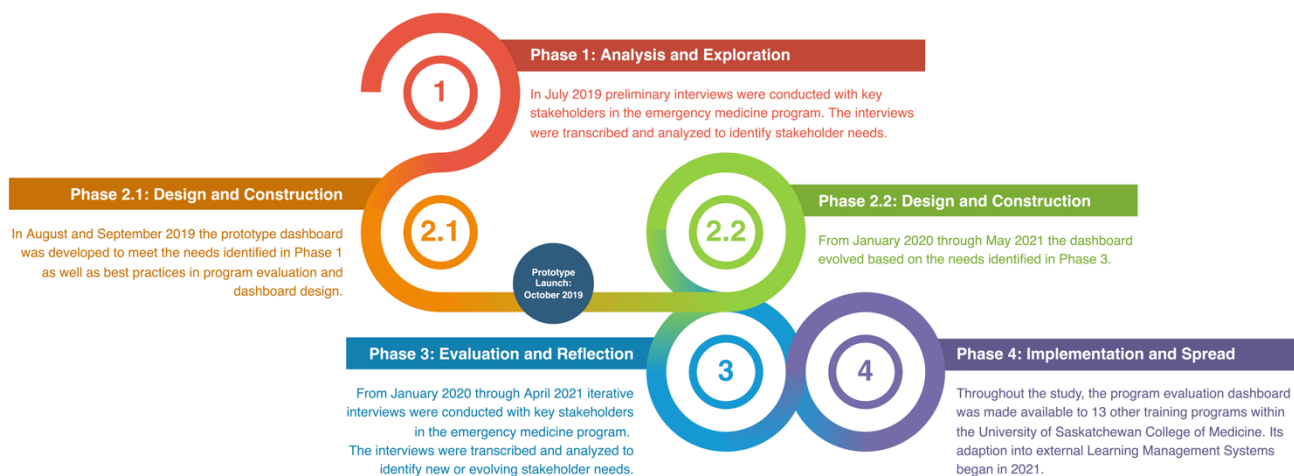
Figure 1. Overview of the four phases of the design-based research methodology.

Data analysis

Narrative data from the interviews was recorded by the senior author (BT) using handheld recording device for in-person interviews and the native Zoom (Zoom Video Recording, San Jose, USA) recording device for virtual interviews. The recordings were transcribed by the University of Saskatchewan's Social Sciences Research Laboratories then thematically analyzed to identify the core needs for the use of CBME assessment data for program evaluation. Dashboard elements (data, analytics, and visualizations) were designed to meet these needs and spurred discussion at subsequent interviews regarding the optimal presentation of the data.

The qualitative analysis was conducted using a constant comparative method.[42] Following the first set of interviews, two authors (YY and RC) independently developed codebooks with representative quotes for each code. They then met and amalgamated their codebooks by adding, modifying, and removing codes on a consensus basis. One author (YY) compiled the codes into a preliminary framework of program evaluation needs. Following each subsequent interview, the same authors coded the data and refined the thematic framework while selecting representative quotes for each need. BT reviewed all the transcripts, codes, and the framework intermittently to ensure that they were comprehensive and representative of the data. TMC reviewed key quotes, codes, and the framework. Both provided additional suggestions to refine the thematic framework. BT liaised directly with the programming team (SW and VB) throughout the analysis to prioritize updates to the program evaluation dashboard. The resulting thematic framework was described using representative quotes as well as images of the dashboard elements mapped to each theme.

The investigators responsible for coding and thematic analysis considered their own positionality and its potential impact on their data interpretation throughout the coding process. RC is an EM resident within the program with an interest in medical education. YY is an external medical education research fellow with expertise in website development and data visualization. TMC is an EM physician with extensive qualitative research experience who has served on the Competence Committee of her own institution's EM training program. BT is an EM physician and was a Residency Program Committee member during the period of study. He previously served as the Program Director, CBD Lead, and Competence Committee Chair of the residency program. We acknowledge that the involvement of two of the coding investigators with the residency program could impact their interpretations of the data. They were mindful of this and its potential impact on their coding throughout the process and their perspectives were balanced by the participation of two external investigators (YY and TMC) in the analysis. Disagreements were discussed until consensus was reached. Other members of our research team included two computer science master's students (VB and SW), a computer science professor (DM), and the EM program director (RW).

Participant checks occurred in two ways. First, each of the interview participants was asked to review the final thematic analysis and provide feedback on any ideas they felt were missing. Second, most participants were interviewed multiple times during the development

process and had the opportunity to provide feedback when the dashboard elements did not meet their needs.

We reported the results of our qualitative analysis in compliance with reporting standards for qualitative research.[43,44]

Data management and dashboard programming

All EPA assessment data for our residency program were entered by faculty into the Royal College of Physicians and Surgeons Mainport ePortfolio (Ottawa, ON). The data were then exported and uploaded to the dashboard each Monday by the EM program administrative assistant. They also updated contextual non-EPA information (e.g., usernames and roles, resident rotation schedules, stages of training) within the dashboard as needed. During the upload process, an automated process reformatted EPA data to tag it with the rotation each resident was on when each EPA was completed. All dashboard data was stored on a secure server in the Department of Computer Science at the University of Saskatchewan.

The dashboard was developed on a distributed web architecture with three components. They included a database server to securely hold the data, a web server for hosting the website, and a back end server to authenticate users and perform CRUD (create, read, update, and delete) operations[20] based upon the privileges granted to each type of user. This allowed each of these parts to be updated independently, facilitating rapid prototyping. A front-end layer of the dashboard visualized the design elements described within this study. These visualizations were rendered in a scale and transform invariant Scalable Vector Graphics (SVG) format that make the user experience consistent across various screen sizes and orientations. Logging into the dashboard required authentication

through the University of Saskatchewan's Central Authentication Service. Access to data was restricted based on pre-assigned user roles. The dashboard source code was published on GitHub[41] under an open access license to allow it to be used by other institutions. There are no plans to commercialize the dashboard.

Results

Four main themes emerged from thematic analysis including: 'tracking changes in overall assessment metrics,' 'comparing metrics to the assessment plan,' 'evaluating rotation performance,' and 'engagement with the assessment metrics.' There were 16 sub-themes. Appendix A outlines the themes along with descriptive participant quotations and their related figures. Video 1 (https://youtu.be/7HCtCrun8-I) is a screencast of the dashboard which outlines the four themes through a demonstration of how users interact with each dashboard element.

1. Tracking changes in overall assessment metrics

Participants requested a broad overview of EPA analytics outlining how EPAs are being completed within the program and how that compares to prior years. Figure 2 depicts overall analysis of EPA metrics from all residents in a single program during the selected academic year. Metrics include the number of residents with records (blue box), total EPAs observed (purple box), percentages of EPAs that expired after not being completed for four weeks (green box), average EPA entrustment score from the 5-point entrustment score (blue box), and average word counts in EPA assessments feedback data (red box). The proportion of EPAs rated at each level of the entrustment score is also presented as a pie chart with a breakdown of each rating.
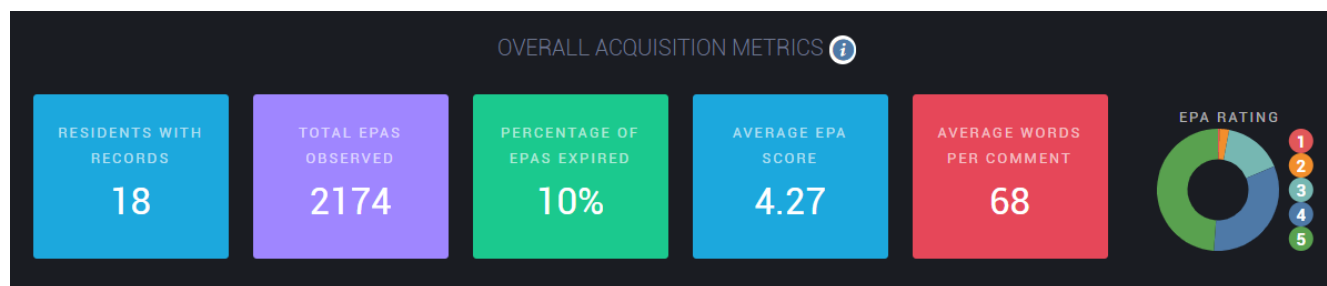

Figure 2. Metrics of EPA completion for the University of Saskatchewan emergency medicine program in the 2020-21 academic year.

Participants required historical metrics to contextualize these values and provide insight into how they were changing in response to faculty development or program changes. Figure 3 presents four visuals (a-d) that display data from multiple academic years:

- Figure 3a is a bar chart presenting the total number of EPAs completed (green bars) and expired (purple bars) per resident. For instance, 2018-2019 and 2019-2020 EPA acquisition and expiration numbers are very similar. As these visuals were captured in April 2021, the 2020-2021 academic year does not have as many completed or expired EPAs.

- Figure 3b is a stack chart displaying the proportion of EPAs rated at each level of the entrustment score from red (1/5) through green (5/5). Overall, the ratings have been quite consistent year-over-year with few EPAs scored at the 1 or 2 level. The only discordant year is 2017-18 when the new assessment system was being piloted.

- Figure 3c is a line graph comparing the number of EPAs completed in each month. In this case, the red (2020-21) trend line drops below orange (2019-20) in January. This coincided with our program's introduction of a non-EPA assessment form.

- Figure 3d is a bar chart contrasting the number of words per comment. Other than the 2017-18 pilot year, the average number of words contained within each EPAs narrative assessment have increased slightly. As described in Appendix A, word count was thought to be a useful but imperfect metric of narrative comment quality.

Participants also requested metrics describing the length of time between the initiation of an EPA to its completion and the proportion of EPAs that were initiated by residents versus faculty. However, they could not be incorporated because our learning management system does not export this data.

2. Comparing metrics to the assessment plan

Beyond knowing the number and characteristics of the EPAs that were being completed within their programs, the participants desired a metric to track the fidelity of their implementation of CBD that compared their program's EPA completion with EPA requirements developed by the Royal College of Physicians and Surgeons of Canada. To meet this need, we developed a 'divergence ratio' that could be calculated for each EPA to quantify its over- or under-representation relative to the prescribed assessment program. For example, for a given transition to discipline (TTD) EPA referred to as TTDx, this metric would be calculated as follows:

$$100\% \; x \left(\frac{\frac{Number\ of\ TTDx\ EPAs\ completed}{Number\ of\ TTD\ EPAs\ completed}}{\frac{Number\ of\ TTDx\ EPAs\ required}{Number\ of\ TTD\ EPAs\ required}} \right)$$
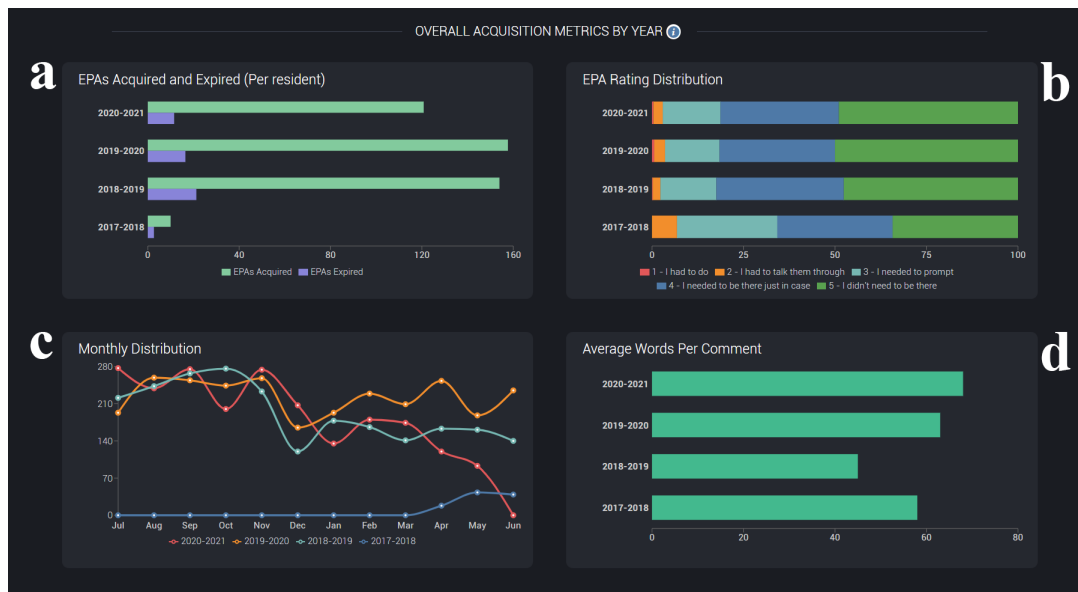

Figure 3. Visualizations of EPA assessment metrics from the University of Saskatchewan emergency medicine program from 2017-2021.

If an assessment program required 10 TTD1 EPAs and 40 TTD EPAs overall and the residents collectively completed 50 TTD1 EPAs and 100 TTD EPAs, the ratio of TTD1/TTD EPAs completed (50/100 = 0.5) divided by the ratio of TTD1/TTD EPAs required by the assessment plan (10/40 = 0.25) would be 200%, signifying that TTD1 was over-represented relative to the other TTD EPAs. As a result of the over-representation of this EPA, the other EPAs are likely to be under-represented.

Figure 4 presents this ratio for each of the EPAs in our EM program graphically to visually demonstrate the degree to which each EPA diverges (is over- or under-represented) in the assessment data relative to the assessment plan. This visual can be displayed for a selected academic year or the trailing 12 months. Using the trailing 12 months ensures that at the beginning of an academic year there is always enough recent data to produce a meaningful visualization. The bars are colored to indicate the degree of divergence from the assessment plan: EPAs completed at >125% of the expected rate are colored light green, those within 75 and 125% of the expected rate are colored green, those <75% of the expected rate are yellow, and those <25% of the

expected rate are colored red. Initially, both over—and under-represented EPAs were colored red to represent their divergence, but this was modified to a lighter green to acknowledge that getting too many EPAs completed is not negative. A lighter green was chosen for these EPAs because their over-representation is an opportunity costs that may prevent the completion of other EPAs.

3. Evaluating rotation performance

Our EM program functions on a rotation-based model where residents work in multiple medical specialties, each of which provide an opportunity to have to one or more EPAs assessed.[10] Participants sought insight into when and how these EPAs were being completed on each rotation.

Figure 5c is a bar chart that presents the average number of EPAs completed per rotation by residents from our program. For each academic year, the value for each rotation is calculated by dividing the total number of EPAs completed on a rotation by the number of residents that have completed that rotation. This allows participants to quickly identify rotations that are not contributing to the assessment program as well as those that are completing more assessments than anticipated.
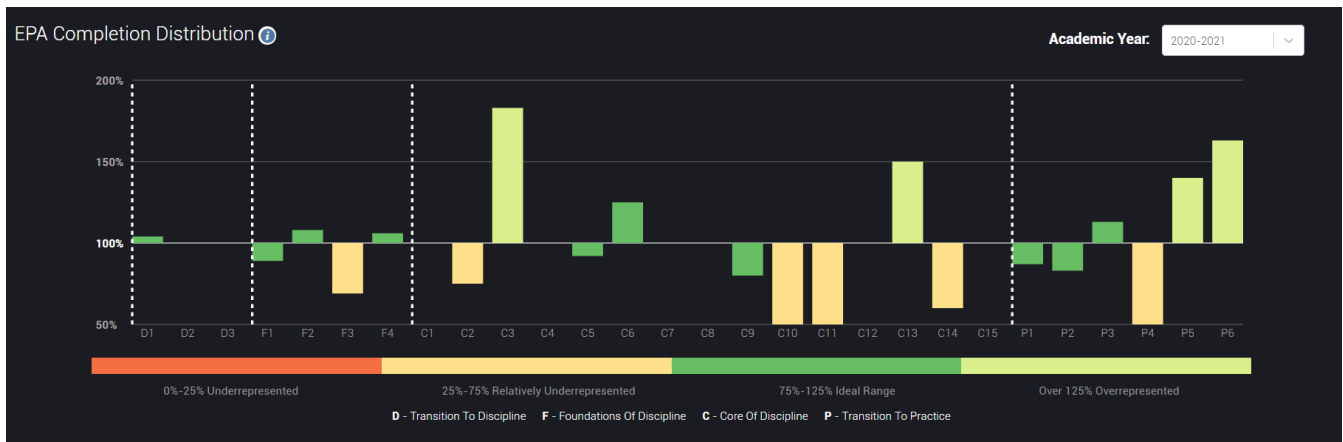

Figure 4. Bar chart visualizing the degree of divergence of the completed assessments from the assessment plan for the University of Saskatchewan emergency medicine program in the 2020-21 academic year.
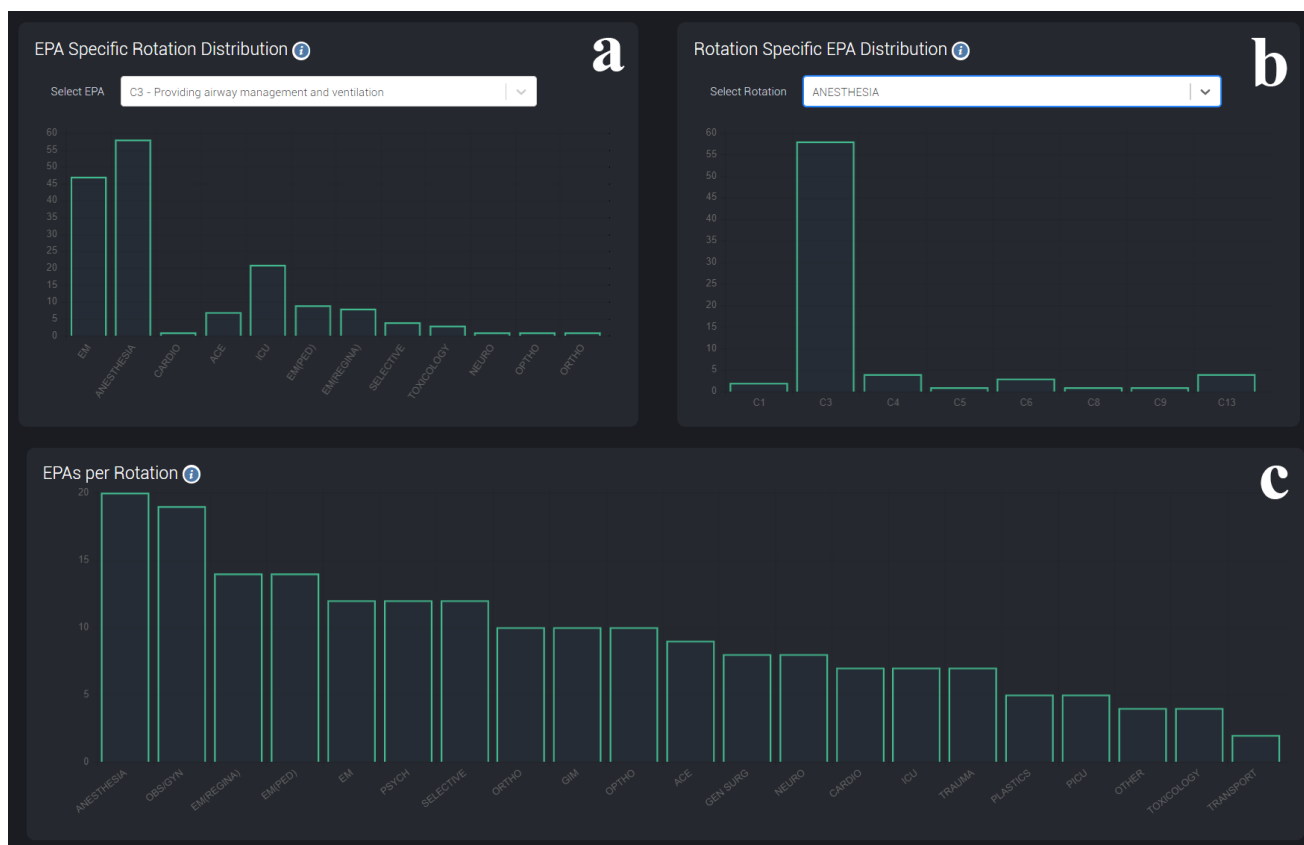
Figure 5. Bar graphs demonstrating the rotations where each EPA is being completed (a) and the EPAs that are being completed on each rotation (b), and the average number of EPAs completed per resident that completed each rotation (c) within the University of Saskatchewan.

Figure 5a and 5b are interactive graphs allow EPA completion and rotation performance to be explored in greater depth. Figure 5a allows users to select a specific EPA and display how frequently it is being completed on each rotation. Conversely, Figure 5b displays the number of EPAs that are being completed while residents are on the selected rotation. These graphs could be used to investigate why an EPA is under-represented or to identify which rotations hard-to-assess EPAs are being completed.

4. Engagement with the assessment metrics

The last theme that emerged related to participants' engagement with the data. Participants wanted the dashboard to be interactive and responsive so that they could efficiently work through their data (see Video 1). Additionally, the continuous availability of data updated in near real-time was helpful to guide ongoing decisions. One final feature that we have not yet been able to incorporate into the dashboard was the development of alerts that would notify the program leadership when metrics were flagged as abnormal. While technically feasible, little literature is available to guide the development of

appropriate metrics to flag and the investigation of this specific question was beyond the scope of our project.

Discussion

Similar to our previous work,[19,20,30] we utilized a design-based research methodology to identify program evaluation needs for data, analytics, and visualizations of CBME assessment data and created a dashboard containing these elements. This approach provides a framework of program evaluation needs for CBME programs. The primary difference in the presentation of data in this work relative to our previous dashboards is that the data is organized largely by rotation and EPA rather than learner[19,20] or faculty member.[30]

The primary strength of this work is the engagement of program leaders from various levels of our institution with real CBME assessment data to guide the identification of program needs and development of dashboard elements. It is also notable that we have published our work on an open-source license so that it can be accessed and modified for use by other institutions as needed.[41] Our practical approach resulted in a dashboard that is having an

21

impact on other Canadian programs and institutions as they is integrate it into their learning management systems.

One of the major challenges in program evaluation within medical education is the varying degrees of fidelity with which an educational program can be implemented.[14] While CBD programs use a shared program of assessment,[9,17] higher or lower levels of fidelity may have a significant impact on trainee outcomes and the ultimate impact of CBME as an intervention. The need for a visualization that quantifies the degree of fidelity of a program's assessments with their stated program of assessment was a major theme that emerged from our analysis and resulted in our new divergence ratio metric.

Unfortunately, the need for ongoing engagement and accessibility of assessment data is, in our experience, not being met at many postgraduate programs. The program evaluation dashboard that we have developed provides dynamic visualizations of data in a near-to-real-time fashion. Challenges in the delivery of the program can therefore be identified and addressed throughout the academic year, rather than at the end. For example, our EM program occasionally highlights under-represented EPAs for residents and faculty to increase awareness and completion of these assessments, which can result in better identification and attempts on these EPAs of interest. More advanced protocols or procedures with greater effect can certainly be adopted,[45] but often interventions will be more successful if the ongoing needs of a particular program are monitored and systematic difficulties are addressed. Similarly, rotations that are not completing EPAs can be identified and engaged as soon as this is identified in the data. This data-driven approach could also inform decisions regarding changes to curriculum.

Limitations

We acknowledge the limitations of this work. For one, it was conducted within a single program at one institution that may have peculiarities that limit its generalizability. It is possible that if our research was conducted in a program with a larger number of possible participants additional themes would have emerged. Further, the involvement of investigators that are closely involved with the training program could have impacted the analysis in unknown ways, although we attempted to ameliorate this through the inclusion of external collaborators.

It is notable that there was some data that we could not incorporate into our dashboard. For example, we identified several dashboard elements that we could not create because we could not access the required data (e.g., time to completion of an EPA and the proportion of EPAs that were resident versus faculty generated). Further, because the focus of our work was on the use of EPA data within the program evaluation of CBME programs, other elements of the program (e.g., activities such as academic teaching sessions, rotation and teaching evaluations) have not been incorporated into the analysis or dashboard.

We acknowledge that the creation of a program evaluation dashboard that meets the needs that we have identified does not address all elements of program evaluation. Mohanna et al. describe program evaluation broadly as "a systematic approach to the collection, analysis, and interpretation of information about any aspect of the conceptualization, design, implementation, and utility of educational programmes."[46] While the described dashboard is a systematic way to collect, analyze, and interpret assessment information about a program's design and implementation, it does not incorporate an evaluation of the utility or outcomes of the educational program. These facets of program evaluation require learning experiences to be tied to outcomes within the context of a program evaluation model (e.g. logic models, Context-Input-Process-Product (CIPP) studies, Kirkpatrick's model, experimental or quasi-experimental models).[47] While the dashboard does not incorporate outcome data or explicitly follow these frameworks, it does present data that would support the use of these models. For example, if one were to conduct a program evaluation exercise using a logic model, the dashboard would provide information on many of the inputs, activities, and outputs of a CBME program that could be tied to downstream clinical and educational outcomes.[13,15]

Lastly, as the implementation of CBME becomes more sophisticated and these outcomes are identified in each medical specialty, it will likely be appropriate to revisit this dashboard to see how such outcomes can be incorporated.[13]

Next steps

This project is a prelude to several other lines of inquiry within a broader program of research. Investigating the impact of the dashboard on our programs was beyond the scope of this investigation, but we hypothesize that our work will support the alignment of learning experiences and the assessment program. Insight into the impact of

CBME could be investigated by quantifying how variability in the implementation of the assessment program affects downstream program outcomes. Further investigation is also needed to identify analytics that warrant the flagging of program leaders.[39] For example, program directors might be interested in being informed if an EPA becomes dramatically under-represented or a rotation is significantly deviating from historical measures. Identifying appropriate alarms or flags and determining how this information should be relayed to program leadership would allow such features to be incorporated into the dashboard. Lastly, further iteration of the dashboard will be required to address diverging and evolving user needs, particularly as it is integrated into widely used Learning Management Systems.

Conclusion

This study identified program evaluation needs that could be met with CBME workplace-based assessment data. Our dashboard and the thematic framework that emerged from interviews with program leaders should inform the development of program evaluation dashboards at other institutions while providing insight into the fidelity of CBME implementations. Future studies should investigate the program evaluation dashboard's effectiveness and contribution to the implementation of CBME and support efforts to quantify the impact of CBME on downstream outcomes.

References

1. Frank JR, Snell LS, Cate O Ten, et al. Competency-based medical education: theory to practice. *Med Teach*. 2010;32(8):638-645. https://doi.org/10.3109/0142159X.2010.501190

2. 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

3. Royal College of Physicians and Surgeons of Canada. *Competence by Design: Reshaping Canadian Medical Education*.; 2014.

4. Englander R, Frank JR, Carraccio C, Sherbino J, Ross S, Snell L. Toward a shared language for competency-based medical education. *Med Teach*. 2017;39(6):582-587. https://doi.org/10.1080/0142159X.2017.1315066

5. Van Loon KA, Driessen EW, Teunissen PW, Scheele F. Experiences with EPAs, potential benefits and pitfalls. *Med Teach*. 2014;36(8):698-702. https://doi.org/10.3109/0142159X.2014.909588

6. Rekman J, Gofton W, Dudek N, Gofton T, Hamstra SJ. Entrustability Scales: outlining their usefulness for competency-based clinical assessment. *Acad Med*. 2015;91(2):1. https://doi.org/10.1097/ACM.0000000000001045

7. Gofton WT, Dudek NL, Wood TJ, Balaa F, Hamstra SJ. The Ottawa Surgical Competency Operating Room Evaluation (O-SCORE): a tool to assess surgical competence. *Acad Med*. 2012;87(10):1401-1407. https://doi.org/10.1097/ACM.0b013e3182677805

8. MacEwan MJ, Dudek NL, Wood TJ, Gofton WT. Continued validation of the O-SCORE (Ottawa Surgical Competency Operating Room Evaluation): use in the simulated environment. *Teach Learn Med*. 2016;28(1):72-79. https://doi.org/10.1080/10401334.2015.1107483

9. Sherbino J, Bandiera G, Doyle K, et al. The competency-based medical education evolution of Canadian emergency medicine specialist training. *Can J Emerg Med*. 2020;22(1):95-102. https://doi.org/10.1017/cem.2019.417

10. Stoneham EJ, Witt L, Paterson QS, Martin LJ, Thoma B. The development of entrustable professional activities reference cards to support the implementation of Competence by Design in emergency medicine. *Can J Emerg Med*. 2019;21(6):803-806. https://doi.org/10.1017/cem.2019.395

11. Royal College of Physicians and Surgeons of Canada. *Entrustable Professional Activities for Emergency Medicine*. Royal College of Physicians and Surgeons of Canada; 2018.

12. Van Melle E, Frank JR, Holmboe ES, Dagnone D, Stockley D, Sherbino J. A core components framework for evaluating implementation of competency-based medical education programs. *Acad Med*. 2019;94(7):1002-1009. https://doi.org/10.1097/ACM.0000000000002743

13. Chan TM, Paterson QS, Hall AK, et al. Outcomes in the age of competency-based medical education: recommendations for emergency medicine training in Canada from the 2019 symposium of academic emergency physicians. *Can J Emerg Med*. 2020;22(2):204-214. https://doi.org/10.1017/cem.2019.491

14. Van Melle E, Hall AK, Schumacher DJ, et al. Capturing outcomes of competency-based medical education: the call and the challenge. *Med Teach*. 2021;43(7). https://doi.org/10.1080/0142159X.2021.1925640

15. Hall AK, Schumacher DJ, Thoma B, et al. Outcomes of competency-based medical education: a taxonomy for shared

language. *Med Teach*. 2021;43(7):788-793. https://doi.org/10.1080/0142159X.2021.1925643

16. Carraccio C, Martini A, Van Melle E, Schumacher DJ. Identifying core components of EPA implementation: a path to knowing if a complex intervention is being implemented as intended. *Acad Med J Assoc Am Med Coll*. Published online March 23, 2021. https://doi.org/10.1097/ACM.0000000000004075

17. Thoma B, Hall AK, Clark K, et al. Evaluation of a national competency-based assessment system in emergency medicine: a CanDREAM study. *J Grad Med Educ*. 2020;12(4):425-434. https://doi.org/10.4300/JGME-D-19-00803.1

18. Hall AKM, Rich J, Dagnone JDM, et al. It's a marathon, not a sprint: rapid evaluation of CBME program implementation. *Acad Med*. 2020;95(5):786-793. https://doi.org/10.1097/ACM.0000000000003040

19. Carey R, Wilson G, Bandi V, et al. Developing a dashboard to meet the needs of residents in a competency-based training program: a design-based research project. *Can Med Educ J*. 2020;11(6):e31-e35. https://doi.org/10.36834/cmej.69682

20. Thoma B, Bandi V, Carey R, et al. Developing a dashboard to meet Competence Committee needs: a design-based research project. *Can Med Educ J*. 2020;11(1):e16-e34. https://doi.org/10.36834/cmej.68903

21. Thoma B, Caretta-Weyer H, Schumacher DJ, et al. Becoming a Deliberately developmental organization: using competency-based assessment data for organizational development. *Med Teach*. 2021;43(7):801-809. https://doi.org/10.1080/0142159X.2021.1925100

22. Stein M, Janetzko H, Lamprecht A, et al. Bring it to the pitch: combining video and movement data to enhance team sport analysis. *IEEE Trans Vis Comput Graph*. 2018;24(1):13-22. https://doi.org/10.1109/TVCG.2017.2745181

23. Teizer J, Cheng T, Fang Y. Location tracking and data visualization technology to advance construction ironworkers' education and training in safety and productivity. *Autom Constr*. 2013;35:53-68. https://doi.org/10.1016/j.autcon.2013.03.004

24. Stadler JG, Donlon K, Siewert JD, Franken T, Lewis NE. Improving the efficiency and ease of healthcare analysis through use of data visualization dashboards. *Big Data*. 2016;4(2):129-135. https://doi.org/10.1089/big.2015.0059

25. Few S. *Information Dashboard Design: The Effective Visual Communication of Data*. 1st Editio. O'Reilly Media; 2006.

26. Ellaway RH, Pusic M V., Galbraith RM, Cameron T. Developing the role of big data and analytics in health professional education. *Med Teach*. 2014;36(3):216-222. https://doi.org/10.3109/0142159X.2014.874553

27. McKenney S, Reeves TC. *Conducting Educational Design Research*. 2nd ed. Routledge; 2019.

28. Reeves TC, Herrington J, Oliver R. design research: a socially responsible approach to instructional technology research in higher education. *J Comput High Educ*. 2005;16(2):97-116.

29. Chen W, Reeves TC. Twelve tips for conducting educational design research in medical education. *Med Teach*. 2020;4(9):980-986. https://doi.org/10.1080/0142159X.2019.1657231

30. Yilmaz Y, Carey R, Chan T, et al. Developing a dashboard for faculty development in competency-based training programs: a design-based research project. *Can Med Educ J*. 2021;12(4):48-64. https://doi.org/10.36834/cmej.72067

31. Boscardin C, Fergus KB, Hellevig B, Hauer KE. Twelve tips to promote successful development of a learner performance dashboard within a medical education program. *Med Teach*. 2017;40(8):1-7. https://doi.org/10.1080/0142159X.2017.1396306

32. Karami M, Langarizadeh M, Fatehi M. Evaluation of effective dashboards: key concepts and criteria. *Open Med Inform J*. 2017;11(1):52-57. https://doi.org/10.2174/1874431101711010052

33. Dolmans DHJM, Tigelaar D. Building bridges between theory and practice in medical education using a design-based research approach: AMEE Guide No. 60. *Med Teach*. 2012;34(1):1-10. https://doi.org/10.3109/0142159X.2011.595437

34. Chen W, Sandars J, Reeves TC. Navigating complexity: the importance of design-based research for faculty development. *Med Teach*. 2021;43(4):475-477. https://doi.org/10.1080/0142159X.2020.1774530

35. Van Melle E, Gruppen L, Holmboe ES, Flynn L, Oandasan I, Frank JR. Using contribution analysis to evaluate competency-based medical education programs: it's all about rigor in thinking. *Acad Med*. 2017;92(6):752-758. https://doi.org/10.1097/ACM.0000000000001479

36. Oandasan I, Martin L, McGuire M, Zorzi R. Twelve tips for improvement-oriented evaluation of competency-based medical education. *Med Teach*. 2020;42(3):272-277. https://doi.org/10.1080/0142159X.2018.1552783

37. Lang C, Siemens G, Wise A, Gasevic D. *Handbook of Learning Analytics: First Edition*. Society for Learning Analytics Research.

38. Chan T, Sebok-Syer S, Thoma B, Wise A, Sherbino J, Pusic M. Learning analytics in medical education assessment: the past, the present, and the future. *AEM Educ Train*. 2018;2(2):178-187. https://doi.org/10.1002/aet2.10087

39. Thoma B, Ellaway R, Chan TM. From Utopia through Dystopia: charting a course for learning analytics in competency-based medical education. *Acad Med*. 2021;96(7S):S89-S95. https://doi.org/10.1097/ACM.0000000000004092

40. Olmos M, Corrin L. Learning analytics: a case study of the process of design of visualizations. *J Asynchronous Learn Netw*. 2012;16(3):39-49.

41. Bandi V, Shisong W, Thoma B. CBD Dashboard UI. GitHub. Published 2019. https://github.com/kiranbandi/cbd-dashboard-ui [Accessed Oct 1, 2021].

42. Watling CJ, Lingard L. Grounded theory in medical education research: AMEE Guide No. 70. *Med Teach*. 2012;34(10):850-861. https://doi.org/10.3109/0142159X.2012.704439

43. Tong A, Sainsbury P, Craig J. Consolidated criteria for reporting qualitative research (COREQ): a 32-item checklist for interviews and focus groups. *Int J Qual Health Care*. 2007;19(6):349-357. https://doi.org/10.1093/intqhc/mzm042

44. O'Brien BC, Harris IB, Beckman TJ, Reed DA, Cook DA. Standards for reporting qualitative research. *Acad Med*. 2014;89(9):1245-1251. https://doi.org/10.1097/ACM.0000000000000388

45. Sample S, Rimawi HA, Bérczi B, Chorley A, Pardhan A, Chan TM. Seeing potential opportunities for teaching (SPOT): Evaluating a bundle of interventions to augment entrustable professional activity acquisition. *AEM Educ Train*. 2021;5(4):e10631. https://doi.org/10.1002/aet2.10631

46. Mohanna K, Cottrell E, Wall D, Chambers R. *Teaching Made Easy: A Manual for Health Professionals*. 3rd ed. Routledge & CRC Press; 2011.

47. Frye AW, Hemmer PA. Program evaluation models and related theories: AMEE guide no. 67. *Med Teach*. 2012;34(5):288-299. https://doi.org/10.3109/0142159X.2012.668637

Appendix A. Thematic analysis of program evaluation needs and the dashboard elements developed to address them.

Main themes and sub-themes	Sample Participant Quotation(s)
1. Tracking changes in overall assessment metrics Figures 2, 3a-d	The--yeah, in terms of acquired and expired, that's nice data to have and see where we're doing as a program and gets me a sense of nice piece of data to go to a department meeting or talk to faculty about and say, "Look, as a group, we're not doing as good a job of this," and then, that can give me some, likely, some support at higher levels.
1.1 Program-wide EPA acquisition Figure 2	super helpful 'cause we've already with our curriculum blueprint figured out where the EPAs should be getting filled out. So this is a great way to look at it and see if there's EPAs that are not getting filled out that should be and then trying to sort that out.
1.1.1 Annual EPA completion Figure 3a	Is there a way we can overlay year to year? 'Cause that would be--that's really what we wanna see, what's changing like what interventions we may have provided and what's working and what's not working.
1.1.2 Monthly EPA completion Figure 3c	Monthly distribution, yeah, that's good, just to make sure it's across the board this is just integrated with our practice and not just sorta a prior to performance review exercise and then. John4 I like the monthly distribution as well. Like December's obviously expected just because I know people take holidays and things like that but if we see things kind of hop off towards the end of the year we can do extra prompting to the residents that like we see you guys are really good in July but then you drop off later in the year. We need to keep that consistent and things like that, and just, kind of see if it's specific rotations during those times versus if it's just more kind of resident behavior.
1.1.3 Modifiers of EPA completion	I would say honestly just for the year. Like the average per resident honestly because we have more residents this year so obviously our EPA count's gonna go up… Yeah. And just at the end of year and say like how many EPAs did our residents get? And it's like okay they're averaging this many per month or this many for the year, and this year they're at.
1.2 Narrative Feedback Quality Figure 2, 3d	What is the expectation for the timeliness of it? Or what should be the content of an EPA and what makes a good EPA? What makes an EPA great, what the residents perceive as being valuable? What does a program or CBD committee perceive to be valuable? So, as you do those quality metrics, maybe get into some of that content, it'd be useful and you can say well, could this be—'cause we're already being assessed on performance and teaching, right, but the metric's very crude… So, I think there's an opportunity to provide that feedback, both to faculty and, at some point, to, obviously, the program and CBD committee, but also to Department Head 'cause department heads—my last assessment, I'll share with you and it could be part of this recording. My last assessment, pretty well every assessment, I think the teaching, the quality of teaching, is not really discussed at all. It's just glanced over.
1.2.1 Word Count Figure 2, 3d	I like the feedback word count. That's good too because that helps I think people strive for it; it can be done. And if you go down, what all did you have down at the bottom there? Oh yeah this. This, like you say this is a little harder to look at because not every program's consistently uploading everything. Like I suspect <program> hasn't uploaded anything for a while, but. But the top I think is very useful.
1.2.2 Quality Evaluation	Obviously we talked before about the quality, the narrative comments beyond word count if that were possible. And then the correlation between entrustment score and comment, if it makes sense, if there was some way to measure that and give people feedback on it, that would be useful.
1.3 Expired EPAs Figure 2, 3a	So I have to admit, you know again going back to <specialty 1>, the expired is very helpful. Because residents will often say "oh they just all expired, we didn't get any" well it's like, less than 20% is not bad. Because there are gonna be some. So I think this is a very nice visual.
1.4 Distribution of EPA entrustment ratings Figure 2, 3b	Probably not related to the program, itself, but we have pretty significant either leniency bias or with how we entrust things or a substantial proportion of the work that we do is relatively easy to do and capturing the stuff that's harder to do and harder to entrust is less frequent. I'm not sure which of those two phenomena is at play, but the fact that the overwhelming majority of our EPAs--what is it? … like 80% of our EPAs or more, are scores of four and five is, it worries me a bit, but good information to have.
1.5 Time to completion of EPAs	Because if I notice one program is always doing them 14 days later that obviously needs some education because that's just not, overly helpful. And if we have one program that's always getting them done within 24 hours we need to figure out what they're doing right.
1.6 Resident versus faculty generation of EPAs	Yeah, the other one I'd like to see, I think I've talked to you about before, is resident generated EPAs versus faculty generated EPAs. I'd like to capture that data and I think that'd be useful in a program evaluation level, just to see are the residents carrying CBD for the program or is the faculty carrying CBD and how's this working.
2. Comparing metrics to the assessment plan Figure 4	It is useful 'cause it's like a nice overview of which EPAs are not being assessed. And then we can then do a deeper exploration such as with your EPA distribution and rotation-specific distribution, it can inform how you use this. Because to tick through this and look at every single one is not gonna be really meaningful. It's unnecessary, right? But to find an EPA, there might be three or four of them, they're like, "Hmm we're not

	getting a lot of those this year." Let's look at those four and let's dive deep into those four and figure out what our challenges and solutions could be.
	It is useful to see it year over year. It kinda talks about your implementation and then sustainability versus fatigue, to see if whatever orientation efforts you did with your faculty are being sustained or further efforts are needed to make sure that's it's done. So, yeah, it's quite useful to see year to year.
2.1 Underrepresented EPAs Figure 4	No, I think that's actually quite—it's very useful because there may be lots of reasons why a certain EPA is not being delivered, but there may be some structural reasons. Maybe the rotations underperforming 'cause resources should be—whatever the case is, right? So, maybe that's an appropriate sort of—the particular rotation should move to a different education experience. So, I really like the graphs 'cause you can see that C3s, and fours, and fives, and sixes, some of these, in that program, were obviously not completed. So, it's really worth asking why that is and find out reasons. Yeah, that's very useful.
2.2 Overrepresented EPAs Figure 4	I would just wanna make sure, with this data, that if we were over target on every EPA, that we weren't sort of punishing ourselves to say that, well, they're not balanced, 'cause I don't really care if they're unbalanced, as long as they get the minimum target for each of them. So, if they needed 10 resus, but they got 15 and they needed 20 undifferentiated patients and they got 30 and they needed contributing to the teamwork which was ten and they got ten, then to me, that's perfect.
3. Evaluating rotation performance Figures 5a-c	I think what program directors have been asking for is the ability to see where EPAs are getting done. So like from a program evaluation, like an individual discipline [inaudible, 00:13:21] I think this would be very helpful. 'Cause right now they can't really see that. And so you're right, we're sort of trying to go through a multi-curriculum map and make sure it's mapped and then this is sorta your check to see does it actually do what you want it to on the ground.
3.1 Underperforming Rotations Figures 5a-c	And that we can better plan for as we need to adjust the schedule going through. And then EPA count per rotation so that's more helpful for me from the CBD lead perspective I would say, just because if we hit a block like where we've had three residents on <rotation> in the last three months and I filter it out and we have no EPAs completed, that's helpful for me if I have to send a prompt to any of the faculty on <rotation> or the rotation lead on <rotation> and just be like, 'hey this is what we've noticed over the last couple of months, we've had a few trainees rotate, we've had some expiry with your faculty's EPA as well. Could we'- we'll just send out this email prompting them just to fill out EPAs.
3.2 Overperforming Rotations Figures 5a-c	Cause we have targets for how many we wanna get per rotation and to look at… see oh, <program>? On average, they're getting eight, right? Which is exactly what we're hoping for or we're hoping for four to eight. Same with <rotation 1> and same with <rotation 2> and they're sort of getting exactly those numbers. So, that's a nice marker of are we hitting where we want. So, off-service are sort of that four to eight range, great and for our more core rotations, we obviously wanna see--our <program> rotations, we wanna be seeing that 14 to 20 range which is exactly what we're seeing and then, things like <rotation 3> and <rotation 4>, they're punching above their weight class, so this would be a nice thing, as part of our six month review, we identify teachers who are sort of over-performing and not necessarily getting a formal award, but just a quick email with a CC to their department head saying, "Just wanted to let you know the residents thought you were a great teacher in the last six months and we really appreciate what you're doing for our program," this would be a nice little feather in the cap to email the <rotation 4> program director, department head, and just say, "We did our six month review of our rotations, when our residents rotate through your service, they're getting an average of 12 EPAs observed per block which is higher than our target of eight. Just wanted to say thank you so much for your commitment to educating our residents," and that sort of stuff goes a long way for keeping up relationships. Then, if we saw one that was sort of under-performing, not the opposite of an email saying, "You suck," but trying to explore why that might be happening.
4. Engagement with the assessment metrics	No, I think it's fine to share that data. I think it would potentially be helpful for like rotations where, or specialties where it's like you rotate through this but you're not entirely sure what EPAs you're gonna have done. You might just use like a shift in counter card because your residents are off-service and you don't necessarily know what to do. But there's actually EPAs that the program fills out for their own residents that their faculty are filling out and they're getting a reliable amount of exposures that are similar, then maybe hey you can actually target this EPA when you're on a rotation. So thinking of it from that side of things. In terms of it actually leaving to meaningful change I think it's really a matter of getting to the people that are on the ground. So if it can be shared with other programs and those programs, like the leads or the representatives from them are able to engage with their faculty and actually use the data to change behavior, I think that's helpful. But I think it's tricky from that sense. But I wouldn't be opposed to sharing it in any way.
4.1 Availability	This is what I love about it is that this helps the CCC or the RPC make real-time decisions on where learning experiencing should be provided. And they don't have to wait for one year of data to come up and get feedback from residents who may or may not speak. So this is truly awesome!
4.2 Notifications of abnormal data	Only if there's an egregious issue, which of I don't know if [inaudible] or not, but to make other things for that, that information and things like that, they will not show up here. But, the other things that require immediate action should be sent right away, otherwise three months.