



Developing a dashboard for faculty development in competency-based training programs: a design-based research project

Élaboration d'un tableau de bord pour la formation professorale dans les programmes de formation axés sur les compétences : projet de recherche orientée par la conception

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

Volume 12, Number 4, 2021

URI: <https://id.erudit.org/iderudit/1082650ar>

DOI: <https://doi.org/10.36834/cmej.72067>

[See table of contents](#)

Publisher(s)

Canadian Medical Education Journal

ISSN

1923-1202 (digital)

[Explore this journal](#)

Cite this article

Yilmaz, Y., Carey, R., Chan, T., Bandi, V., Wang, S., Woods, R., Mondal, D. & Thoma, B. (2021). Developing a dashboard for faculty development in competency-based training programs: a design-based research project. *Canadian Medical Education Journal / Revue canadienne de l'éducation médicale*, 12(4), 48–64. <https://doi.org/10.36834/cmej.72067>

Article abstract

Background: Canadian specialist residency training programs are implementing a form of competency-based medical education (CBME) that requires frequent assessments of entrustable professional activities (EPAs). Faculty struggle to provide helpful feedback and assign appropriate entrustment scores. CBME faculty development initiatives rarely incorporate teaching metrics. Dashboards could be used to visualize faculty assessment data to support faculty development.

Methods: Using a design-based research process, we identified faculty development needs related to CBME assessments and designed a dashboard containing elements (data, analytics, and visualizations) meeting these needs. Data was collected within the emergency medicine residency program at the University of Saskatchewan through interviews with program leaders, faculty development experts, and faculty participating in development sessions. Two investigators thematically analyzed interview transcripts to identify faculty needs that were audited by a third investigator. The needs were described using representative quotes and the dashboard elements designed to address them.

Results: Between July 1, 2019 and December 11, 2020 we conducted 15 interviews with nine participants (two program leaders, three faculty development experts, and four faculty members). Three needs emerged as themes from the analysis: analysis of assessments, contextualization of assessments, and accessible reporting. We addressed these needs by designing an accessible dashboard to present contextualized quantitative and narrative assessment data for each faculty member.

Conclusions: We identified faculty development needs related to EPA assessments and designed dashboard elements to meet them. The resulting dashboard was used for faculty development sessions. This work will inform the development of CBME assessment dashboards for faculty.

© Yusuf Yilmaz, Robert Carey, Teresa M Chan, Venkat Bandi, Shisong Wang, Robert A Woods, Debajyoti Mondal and Brent Thoma, 2021



This document is protected by copyright law. Use of the services of Érudit (including reproduction) is subject to its terms and conditions, which can be viewed online.

<https://apropos.erudit.org/en/users/policy-on-use/>

Developing a dashboard for faculty development in competency-based training programs: a design-based research project

Élaboration d'un tableau de bord pour la formation professorale dans les programmes de formation axés sur les compétences : projet de recherche orientée par 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 Thoma³

¹Continuing Professional Development Office and McMaster Education Research, Innovation, and Theory (MERIT) Program, McMaster University, Ontario, Canada; ²Department of Medical Education, Ege University, Izmir, Turkey; ³Department of Emergency Medicine, University of Saskatchewan, Saskatchewan, Canada; ⁴Emergency Medicine, Department of Medicine, McMaster University, Ontario, Canada; ⁵Department of Computer Science, University of Saskatchewan, Saskatchewan, Canada

Correspondence to: Dr. Brent Thoma, Room 2646, Box 16 103 Hospital Drive, Saskatoon, SK. S7N 0W8; e-mail: brent.thoma@usask.ca; cell: 1 306 881 0112; Twitter: @Brent_Thoma

Published ahead of issue: June 30, 2021; published: September 14, 2021; CMEJ 2021, 12(4). Available at <http://www.cmej.ca>

© 2021 Yilmaz, Carey, Chan, Bandi, Wang, Woods, Mondal, Thoma; licensee Synergies Partners

<https://doi.org/10.36834/cmej.72067>. This is an Open Journal Systems article distributed under the terms of the Creative Commons Attribution License. (<https://creativecommons.org/licenses/by-nc-nd/4.0>) which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is cited.

Abstract

Background: Canadian specialist residency training programs are implementing a form of competency-based medical education (CBME) that requires frequent assessments of entrustable professional activities (EPAs). Faculty struggle to provide helpful feedback and assign appropriate entrustment scores. CBME faculty development initiatives rarely incorporate teaching metrics. Dashboards could be used to visualize faculty assessment data to support faculty development.

Methods: Using a design-based research process, we identified faculty development needs related to CBME assessments and designed a dashboard containing elements (data, analytics, and visualizations) meeting these needs. Data was collected within the emergency medicine residency program at the University of Saskatchewan through interviews with program leaders, faculty development experts, and faculty participating in development sessions. Two investigators thematically analyzed interview transcripts to identify faculty needs that were audited by a third investigator. The needs were described using representative quotes and the dashboard elements designed to address them.

Results: Between July 1, 2019 and December 11, 2020 we conducted 15 interviews with nine participants (two program leaders, three faculty development experts, and four faculty members). Three needs emerged as themes from the analysis: analysis of assessments, contextualization of assessments, and accessible reporting. We addressed these needs by designing an accessible dashboard to present contextualized quantitative and narrative assessment data for each faculty member.

Conclusions: We identified faculty development needs related to EPA assessments and designed dashboard elements to meet them. The resulting dashboard was used for faculty development sessions. This work will inform the development of CBME assessment dashboards for faculty.

Résumé

Contexte : Les programmes de résidence de spécialité au Canada mettent en œuvre une forme d'éducation axée sur les compétences (EASC) qui exige des évaluations formatives fréquentes des activités professionnelles confiées (APC). Les enseignants ont du mal à fournir une rétroaction utile et à attribuer des notes appropriées au niveau de confiance. Les initiatives de formation professorale des enseignants qui interviennent dans la EASC intègrent rarement leurs données psychométriques. Des tableaux de bord pourraient être utilisés pour visualiser les données d'évaluation du corps professoral afin de soutenir leur perfectionnement.

Méthodes : En utilisant un processus de recherche orientée par la conception, nous avons déterminé les besoins de formation professorale liés aux évaluations dans la EASC et nous avons conçu un tableau de bord contenant des éléments (données, analyses et éléments visuels) pour répondre à ces besoins. Les données ont été recueillies dans le cadre du programme de résidence en médecine d'urgence de l'Université de Saskatchewan par le biais d'entrevues avec les responsables du programme, des experts en formation professorale et les enseignants participant aux séances de formation. Deux chercheurs ont procédé à une analyse thématique des transcriptions d'entrevues afin d'identifier les besoins des enseignants, et un troisième chercheur les a vérifiées. Les besoins ont été décrits à l'aide de citations représentatives et des éléments du tableau de bord conçus pour y répondre.

Résultats : Entre le 1^{er} juillet 2019 et le 11 décembre 2020, nous avons mené 15 entretiens avec neuf participants (deux responsables de programme, trois experts en formation professorale et quatre membres du corps professoral). Trois besoins sont ressortis comme thèmes de l'analyse : l'analyse des évaluations formatives, la contextualisation des évaluations formatives et l'accessibilité des rapports. Pour répondre à ces besoins, nous avons conçu un tableau de bord accessible présentant des données d'évaluation quantitatives et narratives contextualisées pour chaque membre du corps professoral.

Conclusions : Nous avons identifié les besoins de formation professorale liés aux évaluations des APC et conçu les éléments d'un tableau de bord permettant d'y répondre. Le tableau de bord a été utilisé dans des séances de formation professorale. Ce travail orientera la réalisation de tableaux de bord afin de faciliter l'évaluation pour les enseignants dans le cadre de la EASC.

Introduction

The Royal College of Physicians and Surgeons of Canada has committed to the implementation of Competency Based Medical Education (CBME) through the *Competence By Design* model.¹ As part of this programmatic assessment^{2,3} model, faculty observe residents' work⁴ and complete frequent, low-stakes assessments of entrustable professional activities (EPAs) that include an entrustment score⁵⁻⁷ and narrative feedback. This is a substantial departure from historical assessment systems and the transition has been difficult for faculty⁸ who struggle to provide accurate entrustment scores and helpful feedback.^{9,10} Identifying and supporting faculty in these areas is considered a key component of the successful implementation of a CBME assessment program.^{11,12}

In clinical medicine, audit and feedback programs have been used to change physician clinical behaviors in some contexts.¹³ These programs are informed by large amounts of clinical performance data. Unfortunately, faculty developers have traditionally had relatively little data about the effectiveness of their teachers and traditional tools such as trainee-driven faculty evaluations provide little insight into performance in these areas and have numerous other flaws and biases.¹⁴⁻²² These limitations have prevented the application of audit and feedback techniques in faculty development. However, the number of EPA assessments completed by faculty within CBME dwarfs that of traditional assessment programs,²³⁻²⁵ overcoming the aforementioned challenge of insufficient data. These EPA assessments represent a small part of learners' educational experience, but contain important feedback that informs competence committees and learners about their progress.²⁶⁻²⁸ The analysis of this data could also provide insights into the quality of the assessments completed by faculty members and support efforts to improve them.

To be leveraged for faculty development, assessment data must be analyzed and visualized in a way that allows faculty developers to identify struggling faculty members and support their remediation. Analytical and visualization techniques have been developed in other fields (e.g. sport and business) to support leaders/coaches and their employees/athletes.²⁹⁻³¹ Dashboards are a tool containing these elements that present analytics visually.^{32,33} A faculty dashboard could visualize CBME assessment data in a way that would support targeted, data-driven faculty development.

Our previous work used a design-based research process³⁴⁻³⁶ to identify resident²⁶ and competence committee²⁷ needs and created dashboards to meet them. We used a similar process to identify and meet faculty development needs.

Methods

We employed an iterative, design-based research process³⁴⁻³⁶ and followed best practices in dashboard design^{37,38} including the collection and analysis of iterative feedback from stakeholders. We report the results of our qualitative analysis in compliance with reporting standards for qualitative research.^{39,40} Our research methodology was deemed exempt from ethical review by the University of Saskatchewan Research Ethics Board (BEH ID 1655).

Settings and participants

This project was situated within the Royal College of Physicians and Surgeons of Canada Emergency Medicine (EM) residency program at the University of Saskatchewan between July 1, 2019 and December 11, 2020. Our program transitioned to the EPA-based *Competence By Design* program in July of 2018.²⁸ There were 14 full time residents enrolled within the program from then until June 30, 2019, 18 enrolled from July 1-June, 2020, and 17 enrolled from July 1-December 11, 2020. Participants included program leaders, faculty development experts from outside of the Department of EM, and EM faculty who participated in faculty development sessions using the dashboard. All participants were asked via email by the lead author (BT) to participate in interviews via phone or video conference.

Research team

Our research team included an established medical education researcher (BT), an education science postdoctoral fellow (YY), an EM resident (RC), two computer science master's students (VB and SW), a computer science professor (DM), our EM program director (RW), and an external expert in medical education research and assessment (TMC).

Design based research process

The Design-Based Research methodology^{38,41} that we employed aligned with the approach outlined in our previous work on competency committee and resident dashboards.^{26,27} Design-based research is an "authentic, contextually aware, collaborative, theoretically focused, methodologically diverse, practical, iterative, and operation-oriented" process^{34,38} which aims to bridge practice and research in education through the integration

of investigation and intervention.^{34,36,41} Our process included three of the four phases of design-based research.^{34,36,42}

Phase 1. Analysis and exploration

The senior author (BT) reviewed the literature on feedback in faculty development,^{43–45} learning analytics,^{46,47} and data visualization^{32,38,46,48} to generate ideas for the initial iteration of the faculty development dashboard. He then facilitated interviews with the program leaders and faculty development experts.

Phase 2. Design and construction

These initial interviews data were transcribed and qualitatively analyzed to inform the creation of a dashboard prototype. The themes from these sessions were vetted by BT and both reviewed and described to the dashboard designers (VB and SW) who designed a prototype dashboard. The dashboard prototype was released to the program leaders in October of 2019.

Phase 3. Evaluation and reflection

Phases 2 and 3 alternated into the next year. Follow up interviews were held with each of the program leaders and faculty development experts that were interviewed in Phase 1 as well as additional Faculty. Interviews with the program leaders and faculty were intentionally scheduled after these participants had used the dashboard as part of an audit and feedback faculty development session. These sessions included the presentation of the faculty members' dashboards and coaching intended to improve their EPA assessments that was informed by the dashboard elements. 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).

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 license⁴⁹ and is now being used by 13 other programs at the University of Saskatchewan (12 other postgraduate programs and the undergraduate program). Further, it is being adapted for use within an international Learning Management System (Elentra, Elentra Consortium, Kingston, ON). While this demonstrates the generalizability of our findings and dashboard to broader contexts, this phase did not contribute directly to the determination of faculty development needs so we did not further describe it in our results.

Qualitative analysis

Narrative data from the interviews was recorded and transcribed. The data were thematically analyzed to identify the core needs for the use of CBME assessment data for faculty development. Dashboard elements (data, analytics, and visualizations) were designed to meet these needs and spurred discussion at subsequent focus groups regarding the optimal presentation of the data.

The qualitative analysis was conducted using a constant comparative method.⁵⁰ Following the first focus group, 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 faculty development needs. Following each subsequent focus groups, 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. He provided additional suggestions to refine the thematic framework throughout the analysis process. He also liaised directly with the programming team (SW and VB) to prioritize updates to the faculty development 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 research fellow who is not directly involved with the 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, *Competence By Design* 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, however, they aimed to be mindful of this and its potential impact on their coding throughout the process. Notably, YY provided a perspective that was external to the program and institution and interviews were held with stakeholders external to the emergency program to ensure that a broader perspective was considered. The involvement of the internal investigators was helpful as it allowed for rapid

and contextualized coding that reduced delays in the iterative dashboard design process.

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, the program leaders and faculty development experts were interviewed multiple times during the development process and had the opportunity to provide feedback when the dashboard elements did not meet their needs.

Data management and dashboard programming

All EPA assessment data for our residency program was entered by faculty into the Royal College of Physicians and Surgeons Mainport ePortfolio (Ottawa, ON). The data was then exported and uploaded to the dashboard each Monday by the EM program administrative assistant. During the upload process, EPA data was reformatted and tagged with the rotation each resident was in when each EPA was completed. Contextual and non-EPA information (e.g., user names and roles, resident rotation schedules) was entered into the dashboard by the EM program administrative assistant. 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: 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.²⁷ This allowed each of these parts to be updated independently, facilitating rapid prototyping, and allowing the dashboard to be adapted for additional programs. The dashboard renders its visualizations 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 GitHub⁵¹ under an open access license to allow it to be used by other institutions. There are no plans to commercialize the dashboard.

Results

We conducted six interviews with two EM residency program leaders (the Program Director and Associate Program Director), five interviews with three faculty development experts (our institutional *Competence By Design* Lead, Director of Faculty Development, and Postgraduate Medical Education Dean), and four interviews with four faculty members (EM faculty who participated in audit and feedback sessions using the dashboard). The participants included five women and four men. The interviews were loosely structured: each began with a demonstration and explanation of the dashboard elements that were new to the participants followed by a request for the participant to walk through how they would utilize the dashboard and what could be added, removed, or improved. The interviewer also asked follow-up questions to clarify the participants' perspectives. Interviews were held between July 1, 2019 and December 11, 2020.

Through iterative qualitative analysis, our team identified three themes within the interview transcripts: analyses of assessments, contextualization of analyses, and accessible reporting. Table 1 is a table containing the themes, sub-themes mapped to the various data visualizations, and descriptive quotes from the interviewed participants. These results are also described below with figures from the dashboard. Appendix A provides a more granular overview of our analysis with additional quotes. Video 1 demonstrates the dynamic functions of the dashboard.

1. Analyses of assessments

The participants requested analyses of both the quantitative and narrative aspects of their EPA assessments. The graphics visualized in Figure 1 provide an overview of these metrics.

Table 1. Thematic analysis of faculty development needs and the dashboard elements developed to address them

| Theme | Subtheme, Dashboard Element and Quotation |
|--|---|
| 1. Analyses of assessments (Figure 1-2, Video 1) | <p>1.1 Quantitative acquisition metrics</p> <p>Faculty Member 1: <i>"I think I'm still trying to refine it and the thing that influenced my feedback most was my average entrustment score and realizing how far to the right I was. It actually--since then, I've been a little bit more aware of when it is okay, so to speak, to give the not fives and when it is okay for me to dictate which EPAs we're doing which was maybe a big learning point for the residents."</i></p> <p>Program Leader 2: <i>"So nice having the total number of EPAs served and then especially if you filter it by day having it for the period because we usually do a check-in every three months with the competence committee to see who's completing EPAs, who maybe needs some prompting for actually getting their EPAs done or for getting more helpful comments so it's good to have that, of that filter."</i></p> <p>1.2 Narrative acquisition metrics</p> <p>Faculty Development Expert 1: <i>"I would think so because I mean I can certainly see that what the paper is saying, that you know, number of words would have a correlation with the quality of the feedback, but we know people who can speak well and people who can, you know, need more words to say. But ultimately it has to--for me, ultimately it has to be meaningful to the residents and to the CCC to make their decision, right? So if there was any way to capture residents' opinion on this or the CCC's assessment of that feedback would be really helpful. I would think, going forward."</i></p> <p>Faculty Development Expert 2: <i>"I really feel that anytime you just kind of give one number, like an average thing, that only gives limited picture, right? So if there was any way to provide some measure of dispersion whether it's standard deviation or range, whatever else you wanted to provide, like this is the average EPA score but this is their whatever it might be. And same thing for the average words. I think that gives a bit more meaningful information."</i></p> |
| 2. Contextualization of analyses (Figure 1-5, Video 1) | <p>2.1 By peer</p> <p>Faculty Development Expert 3: <i>"I love the comparisons because I think as physicians we're quite competitive and sometimes it helps us to know where we are in the pack; to know just how much time and effort we'd have to put into something to pull up our socks. But I would want a way to help them, and I think you did show me this, where they can track themselves over time to see how well they're improving on specific parameters."</i></p> <p>2.2 By rotation</p> <p>Program Leader 1: <i>"And similarly with faculty we could look at that rotation and say, 'Look, this is the rotation we really need to get this one observed. Can your faculty please target it?'"</i></p> <p>2.3 By time</p> <p>Faculty Member 4: <i>"I think that is helpful. That gives you a little bit more of the background and the ability to compare a certain time frame. It definitely gives you information on how you're doing in terms of EPAs filling out and stuff."</i></p> <p>Program Leader 1: <i>"And then almost like putting a marker, a date marker on here of like your filter could be since you received your evaluation, right? On this day. Like pre-evaluation, post-evaluation trend and comparison."</i></p> <p>2.4 By resident</p> <p>Program Leader 1: <i>"Being able to sort by EPA, by resident you worked with, by the date, and by the rating, I think the EPA and the rating are gonna be more useful. 'Cause you can look and see, okay, when did I give ones and twos? When did I give fours and fives? And how often am I doing that? And then specific if there's an EPA that seems to be challenging for me, then I can focus in on that and see where my score ranges are. That would be useful."</i></p> <p>2.5 By assessment system</p> <p>Faculty Member 4: <i>"So, I'm just imagining something that could kind of show you some semblance of this, but also with the focus of which (EPAs) have you done more and less of relative to everyone else... It just gives you somewhere to focus, I think, too, to help them get that specific learning experience or to really seek it out and be on the lookout for it 'cause there are gonna be rare ones that you could be like, 'Oh, did we see it this shift? So, let's make sure we go take it and take the opportunity to fill out an EPA on it.'"</i></p> |
| 3. Accessible and Clear Reporting (Supplemental Data, Video 1) | <p>3.1 Accessible Reporting</p> <p>Faculty Development Expert 1: <i>"It'd be helpful to have this be downloadable as some sort of a report that could be given to faculty with context or discussed."</i></p> <p>Faculty Member 3: <i>"Can you add that--I know I get my residents, basically what I've had on the dashboard is what I'm pulling up here in front of me, now, but could you add your own metrics that you could also track yourself in the way of, obviously, things you're doing and how you're comparing up to the group? 'Cause that would--it's your data right? This is probably shared with the Program Directors and probably yourself, basically, right? But, that would--yeah, that might be interesting to look at your--almost like your faculty score or whatever you're calling this, basically."</i></p> <p>3.2 Clear Reporting</p> <p>Program Leader 1: <i>"And then, the other is, is it possible, when I mouse over 4.2, it would tell me what 4.2 is? Because when I look at this, I'm gonna--as a faculty, I'm gonna look at this and say, 'Oh geez, I never fill out EPA blah, blah, blah,' right? What is that? And then, I could just mouse over the number 3.8 and see, oh okay, that's--3.8, I forget. It's like tox or obstetrical or whatever it is, but yeah okay, I can start filling more of that one out or whatever else, as opposed to having to go to a separate list. ... Because when I look at this, I'm gonna--as a faculty, I'm gonna look at this and say, 'Oh geez, I never fill out EPA blah, blah, blah,' right? What is that? And then, I could just mouse over the number 3.8 and see, oh okay, that's--3.8, I forget. It's like tox or obstetrical or whatever it is, but yeah okay, I can start filling more of that one out or whatever else, as opposed to having to go to a separate list."</i></p> |
| Legend: EPA: Entrustable Professional Activity, CCC: Competence Committee Chair, CBD: Competence By Design | |

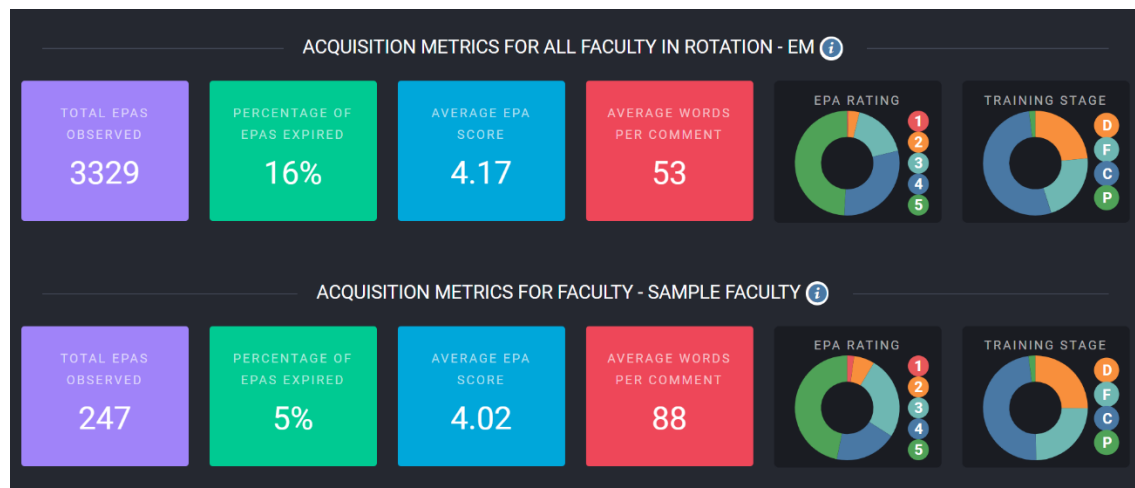


Figure 1. Visual representation of acquisition metrics for all faculty (top row) and the selected faculty member (bottom row)

Quantitative analyses included the number of EPAs completed (purple square), percentage of EPAs that were sent but never completed (green square), average EPA entrustment score (blue square), proportion of ratings at each level of entrustment (pie chart titled 'EPA Rating'), and proportion of EPAs completed for learners at each level of training (pie chart titled 'Training Stage'). Mousing over the EPA rating and training stage visuals opens a pop-up window with the percentage of each slice of the pie charts. There were two quantitative metrics that we were unable to incorporate into the dashboard because they could not be exported from our learning management system: the proportion of EPA assessments initiated by the faculty member (rather than the resident) and the average length of time between the initiation of an EPA assessment and its completion.

Narrative analyses quantified the text submitted with each EPA. The red square contains the average number of words contained in each narrative description. As there is some evidence that this metric correlates with assessment quality in other contexts,⁵² participants were interested in this variable with the caveat that "...you can write a lot of stuff and say very little!" (Faculty Member 3). The participants desired a more robust quality metric, but at this time we are not aware of any that could be automated for visualization. The participants also wanted a way to review the narrative comments that they had submitted. Figure 2 shows how these comments appeared on the dashboard for review. This table can be sorted by column and searched for keywords.

| Summary of EPAs by Sample Faculty | | | | |
|-----------------------------------|---------------|-----|--------|---|
| DATE | RESIDENT NAME | EPA | RATING | FEEDBACK |
| 2021-01-06 | Resident Name | P1 | 5 | Dr X, came on to a busy shift at the end of an afternoon at RUH. There were around 8 patients WTBS. I got stuck calling +++ labs and he was left to begin managing the shift on his own. He picked up the appropriate patients and got workups started on one while seeing a couple of others. I would recommend that in the future he rely more on the PLNs for non-acute psychiatry patients to improve his flow, but he did a great assessment on these patients and was able to disposition them appropriately. With the addition of the rover, the backlog was worked through and despite there being a relatively high volume of arrivals during our shift, he was able to keep the department functioning smoothly. I did relatively little - managing a single handover that we received and the many, many lab follow ups. As I had little to do, I observed Dr X's interactions with many of his patients. He was able to see them efficiently, stay on top of his charts, and manage the department without prompts. |
| 2021-01-06 | Resident Name | C9 | 5 | Dr X, saw a patient with anxiety and some SI. His documentation was excellent. He provided all appropriate supports. The patient's management of interactions with the PLN were good when they presented with some concerns/suggestions. Dr X, had already addressed this and was able to communicate that in a very polite and professional way in a collaborative conversation. All were comfortable with the plan. |

Figure 2. Visual representation of the table containing narrative comments that can be searched and filtered by date, resident, EPA, and entrustment score

2. Contextualization of analyses

The participants requested that the dashboard provide context for the analyses by incorporating analytics on peers, rotations, time, residents, and the assessment system. This provided a better understanding of their relative performance. At the top of the dashboard a drop-down menu allows users to load the dashboard of a particular faculty member. Their assessment data was then contextualized using filters and visualizations.

Another drop-down menu at the top of the dashboard filters the data by *rotation*. This allows individual faculty member's data to be contextualized within the data of faculty completing EPAs on the same rotation. It also facilitates the investigation and remediation of rotation-specific faculty development challenges. For example, if the EM program held a faculty development event for faculty that work on the trauma rotation, they could provide faculty data specific to that rotation. The rotation filter impacts the quantitative and narrative metrics presented at the top of Figure 1 and which faculty are included in Figures 3 and 4.

Components of Figures 1 and 2 also contextualize the selected faculty's data by *resident*. Figure 1 includes the proportion of EPAs completed on residents in different stages of training. This is important as faculty who predominantly work with Transition to Discipline residents

(the first of the four stages of training) are likely to provide feedback that differs from faculty working with Transition to Practice (the last of the four stages of training) residents. Figure 2 allows the filtering and search of the narrative comments by resident to facilitate the review of assessments completed on different types of learners (e.g., searching for residents who are known to be struggling or excelling or by high or low entrustment score ratings to examine how the feedback and entrustment scores vary across learners and contexts).

Figure 3 contextualizes the faculty member's performance among their *peers*. Each chart positions the selected faculty's data (the red bar) in comparison to the metrics of other faculty. The bar charts present the number of EPAs completed (top left), proportion of EPAs that expire prior to completion (top right), average entrustment score (bottom left), and average number of words per narrative comment (bottom right). For program leaders, mousing-over any of the bars displays the faculty member's name and metrics and clicking on any of the bars loads that faculty member's profile. This makes it easy to identify and load the profiles of outlier faculty members on each of these metrics (e.g., faculty with consistently high EPA expiry rates, high or low average entrustment scores, or low numbers of words per comment).

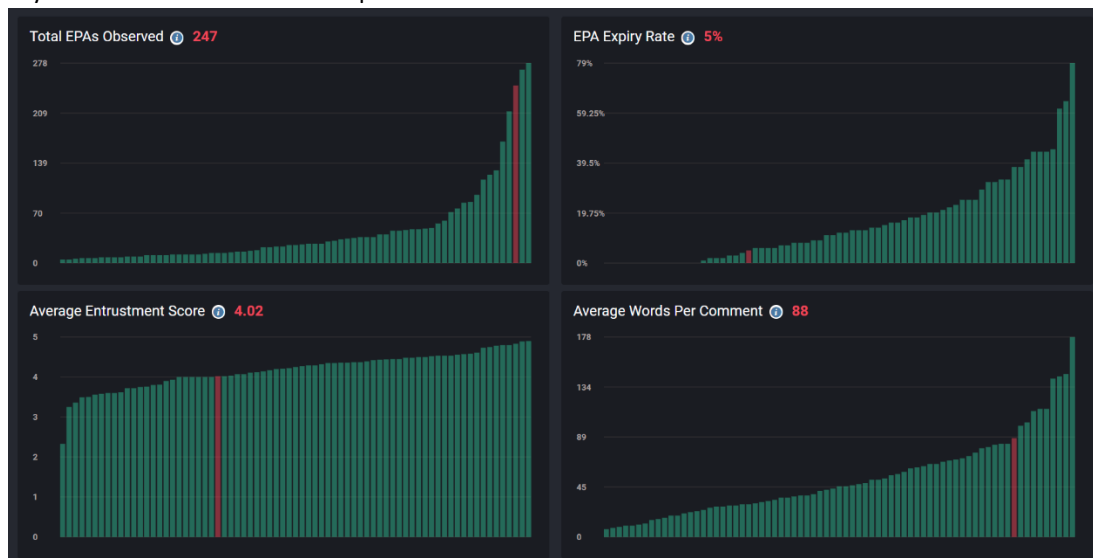


Figure 3. Bar chart of faculty EPA observations with each faculty member represented as a bar and the selected faculty highlighted in red.

Figure 4 presents the same bar charts as Figure 3 with a *time* filter. A date filter at the top of the dashboard allows a date range to be selected that overlays a second bar for the selected time interval on top of the overall value. This allows participants to investigate how faculty members

metrics have changed. For example, if a faculty member with a high EPA expiry rate attended a faculty development initiative, this filter could be used to overlay the expiry rate in the period following this event to see if it had improved.

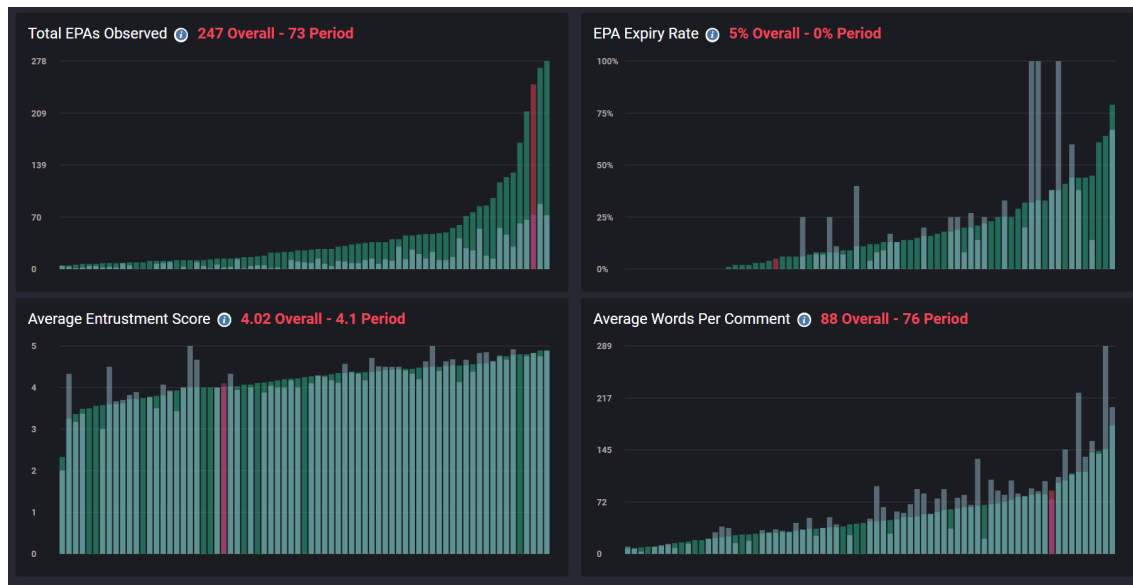


Figure 4. Bar chart of faculty EPA observations with each faculty member represented as a bar with overlaid time-filtered data.

Figure 5 is a spider graph that conveys how frequently the selected faculty member completes each of the EPAs relative to their *peers* (other faculty members) and the requirements of the overall *assessment program*. We developed this visualization in response to a request from a faculty member for information regarding which EPAs were over- or under-represented in their assessment data (e.g., are there some EPAs that they fill out more than needed and others that they do not fill out enough?). The resulting element overlays the representation of each EPA in the data of the selected faculty member (blue) and overall program (purple) both relative to the requirements of the assessment program.

For example, where the number of Transition to Discipline EPA#1 completed by the faculty is D1, the number of all Transition to Discipline EPAs completed by the faculty is D,

the number of Transition to Discipline EPA#1 required by the assessment program is R1, and the number of Transition to Discipline EPAs required by the assessment program is R, the graphed value for that EPA would be calculated as follows:

$$\left(\frac{D1}{D}\right) / \left(\frac{R1}{R}\right)$$

Values for all faculty are calculated using the same formula with D1 and D replaced with values for *all* faculty members. Mousing over the spider graph presents the numerical values for each EPA. Overlaying the data in this way demonstrates how the selected faculty's EPA completion (blue) differs from other faculty (purple) and the requirements of the assessment program. It can be used by faculty and faculty developers to identify and target the completion of underrepresented EPAs.

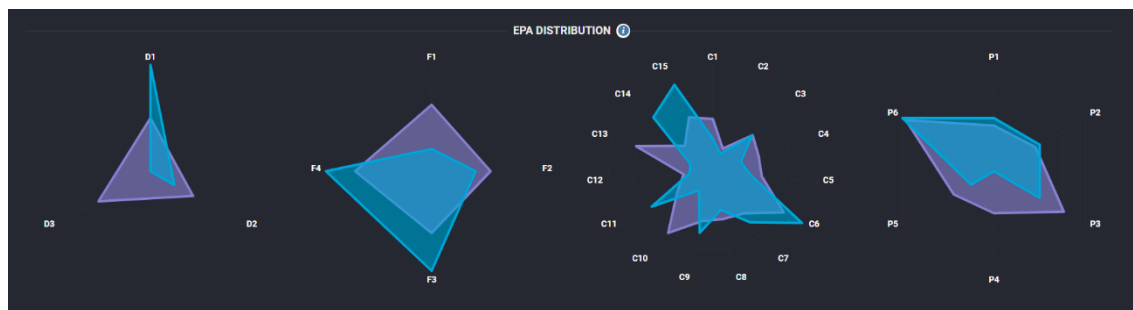


Figure 5. Spider graph visualizing over- and under-completion of EPAs by the selected faculty (blue) and overall program (purple)

3. Accessible and clear reporting

Participants wanted the dashboard to be clear both in its organization and the information that it is presenting. Within the platform, accessibility involved making the dashboard elements easy to navigate and interpret. To meet this need we organized the dashboard in an intuitive way. Filters were placed at the top and were followed moving down the page by broad analytics (Figure 1), comparative data (Figures 3, 4, and 5), and narrative data (Figure 2). To ensure that the purpose of each element was clear, information icons were placed within each element that provided detailed descriptions on mouse-over.

The participants also thought the information should be accessible. They requested the ability to export the information presented in the dashboard for use in faculty development initiatives with individuals or groups. We developed a PDF export containing the currently displayed data and filters to meet this need. Faculty developers can load the dashboard of a specific faculty member, apply the needed filters, and export the associated dashboard elements for download or printing. Notably, as the dashboard contains sensitive identifiable resident assessment data within the narrative assessment table (Figure 2), we also developed a version of the export that excludes this table. This export could be printed and given to the faculty member without concern that this information would be inappropriately lost or disclosed. Supplemental data provides a sample export of a faculty report that excludes the table of narrative comments.

Lastly, provided that the identity of their peers could be concealed, participants were interested in faculty being granted the ability to login and review their contextualized assessment data. We plan to implement this feature in the next update of the dashboard.

Discussion

We utilized a design-based research method to identify faculty development needs for data, analytics, and visualizations of CBME assessment data and created a dashboard to facilitate faculty development. This approach provided a scholarly framework to investigate faculty needs with regards to the utilization of CBME assessment data while also improving our local assessment systems and programs.

Similar to our previous work,^{26,27} we utilized design-based research to identify the needs of faculty members within a CBME system and develop a prototype dashboard to meet

them. The primary difference in the presentation of the assessment data in this work is that the data is organized by individual faculty member rather than an individual learner. This reflects the shift in focus from the learner as the subject of interest to the faculty member and resulted in the creation of entirely new dashboard elements. Whereas the resident and competence committee dashboards *digitalized* the assessment data to support existing processes (resident learning and assessment), the faculty development could be considered a form of *digital transformation* as the data is being used for an entirely new purpose.⁵³

Notably, within this project we used data that was collected for one purpose (resident learning and assessment) for another (faculty development). As considering evidence out of its context can be misleading, it is important to assess its validity in the new context.^{54,55} When seeking validity evidence, we want to ensure that a measure reflects the construct we aim to measure.⁵⁶ In this case, we believe that the use of CBME assessment data to gain insight into faculty assessment behavior is entirely appropriate because we are presenting metrics of their performance in this task. Still, further study will be required to build a validity argument supporting the use of our dashboard and the elements it contains as a faculty development tool. A study demonstrating the efficacy of the audit and feedback process that we have outlined would be an excellent first step but was beyond the scope of this manuscript.

We also caution against the generalization of the dashboard metrics for the evaluation of performance in related areas (e.g., teaching or resident supervision). Given the difficulty of effectively evaluating teaching performance using only learner generated data,^{14–22} the use of these analyses for these purposes is particularly tempting. However, we did not aim to address these needs and feel that other metrics would be better suited to evaluate them. We propose that, like trainees in CBME programs, faculty members should be assessed in a programmatic fashion across a plethora of metrics. The analysis of CBME assessment data should be used *alongside many other faculty performance metrics* to help faculty members to improve. As these metrics are developed it may be reasonable to incorporate them in a dashboard such as the one described here. Ultimately, we feel that providing faculty with feedback on their performance in a way that is analogous to what we are providing to our learners within CBME programs is

important to the development of an organizational culture of feedback⁵³. This approach aligns with a growing body of literature calling for and demonstrating the incorporation of feedback and coaching for faculty as an important approach to faculty development.^{57,58}

Future directions

We anticipate that faculty dashboards will become an important part of CBME programs as they mature. As larger datasets are collected it will be possible to conduct increasingly sophisticated analyses using machine learning algorithms and natural language processing.⁵⁵ These analytics are likely to allow greater insight into faculty assessment behavior and understanding regarding how assessment behavior can be improved.

There were four elements that were requested that could not be implemented due to technical limitations. Two of them (whether the faculty or resident initiated the EPA assessment and how long it took for each EPA assessment to be completed after it was initiated) could not be incorporated because this data is not exported by our learning management system. We hope that these features will be added to the data export in the future so that these needs can be addressed. The third need was for faculty access to their own dashboard, and this will be incorporated into future versions of the dashboard. The final need was for a more sophisticated measure of narrative comment quality than word count, as the participants did not feel that the number of words per comment was an adequate measure of comment quality. Notably, a tool for the assessment of narrative comments in competency-based assessments has already been developed and demonstrated some validity evidence.⁴⁴ In the future, we anticipate that natural language processing algorithms could assess narrative comments for quality using such a tool and report its results on a faculty dashboard along with guidance regarding how the faculty member's assessment feedback could be improved. More advanced still, it may be possible to have such an algorithm analyze comments as they are being entered and provide this guidance to the faculty member at the 'point of assessment.'

While the focus of our work was on CBME assessment data, there is substantial potential for the use of faculty development dashboards that extend beyond this domain. For example, data is also collected on faculty teaching evaluations, research productivity, and clinical performance. A single dashboard encompassing faculty

performance in multiple domains could support coaching and development in all these areas while also providing departments and institutions with metrics that could inform data-driven promotion decisions.

Limitations

There were limitations to our project. First, the data were collected at a single center in Canada and its focus on EM faculty and faculty developers may limit its transferability. However, its use in multiple other specialties and the organizational consistency of *Competence by Design* across Canadian medical specialties suggests that many of our findings will be transferrable. Second, we gathered and analyzed data from three distinct groups (faculty, faculty developers, and program leaders) together. We acknowledge that these groups may have distinct needs that could have been overlooked or misconstrued due to this approach. However, we believe that it was justified given the substantial overlap in their feedback. While some features may have been more desired by one of the groups, none of the groups requested customizations of the dashboard that conflicted with the needs of the other groups. Third, the involvement of investigators who are closely involved in the residency program (e.g., RC is a resident and BT is a faculty member) may have biased the qualitative analysis. We attempted to remediate concerns about this through the inclusion of participants outside of the program and the involvement of an external investigator (YY) in the qualitative analysis.

Conclusions

This project identified multiple needs for the presentation of CBME assessment data to support faculty development within competency-based programs. The resulting dashboard and its supporting thematic framework should inform the development and evolution of faculty development dashboards at other institutions. As our institutions become increasingly sophisticated with their use of data, we anticipate that faculty dashboards will be used to facilitate the data-driven coaching of faculty members within faculty development programs.

Conflicts of Interest: Dr. Yilmaz is the recipient of the TUBITAK Postdoctoral Fellowship grant. The remaining authors have no conflict of interest. There are no plans for the commercialization of the described dashboard. The dashboard code has been published under an open access license.

Funding: This project was supported by a University of Saskatchewan College of Medicine Strategic Grant from the Office of the Vice Dean of Research.

Acknowledgement: The authors would like to acknowledge the University of Saskatchewan Faculty who participated in this research and our Program Administrator, Julie Hunter, for their support of this research.

References

- Royal College of Physicians and Surgeons of Canada. Competence by Design: reshaping Canadian medical education. Ottawa, ON; 2014.
- Frank JR, Snell LS, Cate OT, et al. Competency-based medical education: theory to practice. *Med Teach*. 2010;32(8):638-45. <https://doi.org/10.3109/0142159X.2010.501190>
- Holmboe ES, Sherbino J, Long DM, Swing SR, Frank JR, For the International CBME collaborators. The role of assessment in competency-based medical education. *Med Teach*. 2010;32(8):676-82. <https://doi.org/10.3109/0142159X.2010.500704>
- 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>
- 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>
- 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):1401-7. <https://doi.org/10.1097/ACM.0b013e3182677805>
- 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):72-9. <https://doi.org/10.1080/10401334.2015.1107483>
- Hall AK, Rich J, Dagnone DJ, et al. It's a marathon, not a sprint: rapid evaluation of CBME program implementation. *Acad Med*. 2020 May;95(5):786-793. <https://doi.org/10.1097/ACM.0000000000003040>
- Acai A, Li SA, Sherbino J, Chan TM. Attending emergency physicians' perceptions of a programmatic workplace-based assessment system: the McMaster Modular Assessment Program (McMAP). *Teach Learn Med*. 2019;31(4):434-44. <https://doi.org/10.1080/10401334.2019.1574581>
- Li S-A, Acai A, Sherbino J, Chan TM. The teacher, the assessor, and the patient protector: a conceptual model describing how context interfaces with the supervisory roles of academic emergency physicians. *AEM Educ Train*. 2021;5(1):52-62. <https://doi.org/10.1002/aet2.10431>
- Holmboe ES, Ward DS, Reznick RK, et al. Faculty development in assessment: the missing link in competency-based medical education. *Acad Med*. 2011;86(4):460-7. <https://doi.org/10.1097/ACM.0b013e31820cb2a7>
- Dath D, Lobst W. The importance of faculty development in the transition to competency-based medical education. *Med Teach*. 2010;32(8):683-6. <https://doi.org/10.3109/0142159X.2010.500710>
- Tuti T, Nzinga J, Njoroge M, et al. A systematic review of electronic audit and feedback: intervention effectiveness and use of behaviour change theory. *Implement Sci*. 2017 May 12;12(1):61. <https://doi.org/10.1186/s13012-017-0590-z>
- Steinert Y, Mann K, Centeno A, et al. A systematic review of faculty development initiatives designed to improve teaching effectiveness in medical education: BEME Guide No. 8. *Med Teach*. 2006 Jan 1;28(6):497-526. <https://doi.org/10.1080/01421590600902976>
- Emery CR, Kramer TR, Tian RG. Return to academic standards: a critique of student evaluations of teaching effectiveness. *Qual Assur Educ*. 2003 Jan 1;11(1):37-46. <https://doi.org/10.1108/09684880310462074>
- Hornstein HA. Student evaluations of teaching are an inadequate assessment tool for evaluating faculty performance. Law HFE, editor. *Cogent Educ*. 2017 Jan 1;4(1):1304016. <https://doi.org/10.1080/2331186X.2017.1304016>
- Clayson DE. Student evaluations of teaching: are they related to what students learn?: a meta-analysis and review of the literature. *J Mark Educ*. 2009 Apr 1;31(1):16-30. <https://doi.org/10.1177/0273475308324086>
- Hessler M, Pöpping DM, Hollstein H, et al. Availability of cookies during an academic course session affects evaluation of teaching. *Med Educ*. 2018;52(10):1064-72. <https://doi.org/10.1111/medu.13627>
- Boring A, Dial UMR, Ottoboni K, Stark PB. Student evaluations of teaching (mostly) do not measure teaching effectiveness. *Sci Res*. 2016;(612413):1-36. <https://doi.org/10.14293/S2199-1006.1.SOR-EDU.AETBZC.v1>
- Beran TN, Rokosh JL. Instructors' perspectives on the utility of student ratings of instruction. *Instr Sci*. 2009 Mar 1;37(2):171-84. <https://doi.org/10.1007/s11251-007-9045-2>
- Leone-Perkins M, Schnuth R, Kantner T. Preceptor-student interactions in an ambulatory clerkship: gender differences in student evaluations of teaching. *Teach Learn Med*. 1999 Jul 1;11(3):164-7. <https://doi.org/10.1207/S15328015TL110307>
- Gonyea NE, Osick M, Bradley EG. An Evaluation of the Relationship between Instructor Appearance and College Student Evaluations of Teaching. *J Res Educ*. 2018;28(2):66-92.
- 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):178-87. <https://doi.org/10.1002/aet2.10087>
- 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 Jul 24;12(4):425-34. <https://doi.org/10.4300/JGME-D-19-00803.1>
25. Chan T, Sherbino J, Collaborators for the M. The McMaster Modular Assessment Program (McMAP): a theoretically grounded work-based assessment system for an emergency medicine residency program. *Acad Med*. 2015 Jul;90(7):900-5. <https://doi.org/10.1097/ACM.0000000000000707>
 26. 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-e45. <https://doi.org/10.36834/cmej.69682>
 27. Thoma B, Bandi V, Carey R, et al. Developing a dashboard to meet Competence Committee needs: a design-based research project. *Can J Med Educ*. 2020;11(1):e16-34. <https://doi.org/10.36834/cmej.68903>
 28. Sherbino J, Bandiera G, Doyle K, et al. The competency-based medical education evolution of Canadian emergency medicine specialist training. *Can J Emerg Med*. 2019;1-8. <https://doi.org/10.1017/cem.2019.417>
 29. 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 Jan;24(1):13-22. <https://doi.org/10.1109/TVCG.2017.2745181>
 30. 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 Nov 1;35:53-68. <https://doi.org/10.1016/j.autcon.2013.03.004>
 31. 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 Jun 1;4(2):129-35. <https://doi.org/10.1089/big.2015.0059>
 32. Few S. *Information dashboard design: the effective visual communication of data*. 1st Editio. Cambridge (MA): O'Reilly Media; 2006.
 33. Ellaway RH, Pusic MV, Galbraith RM, Cameron T. Developing the role of big data and analytics in health professional education. *Med Teach*. 2014;36(3):216-22. <https://doi.org/10.3109/0142159X.2014.874553>
 34. McKenney S, Reeves TC. *Conducting Educational Design Research*. 2nd ed. New York: Routledge; 2019. <https://doi.org/10.4324/9781315105642>
 35. 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. <https://doi.org/10.1007/BF02961476>
 36. Chen W, Reeves TC. Twelve tips for conducting educational design research in medical education. *Med Teach*. 2019;0(0):1-7.
 37. Karami M, Langarizadeh M, Fatehi M. Evaluation of Effective Dashboards: Key Concepts and Criteria. *Open Med Inform J*. 2017; <https://doi.org/10.2174/1874431101711010052>
 38. 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>
 39. 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; <https://doi.org/10.1037/t74064-000>
 40. O'Brien BC, Harris IB, Beckman TJ, Reed DA, Cook DA. Standards for reporting qualitative research. *Acad Med*. 2014;89(9):1245-51. <https://doi.org/10.1097/ACM.0000000000000388>
 41. 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>
 42. Chen W, Sandars J, Reeves TC. Navigating complexity: The importance of design-based research for faculty development. *Med Teach*. 2020 Jun 12;1-3.
 43. Warm E, Kelleher M, Kinnear B, Sall D. Feedback on feedback as a faculty development tool. *J Grad Med Educ*. 2018 Jun 1;10(3):354-5. <https://doi.org/10.4300/JGME-D-17-00876.1>
 44. Chan TM, Sebok-Syer SS, Sampson C, Monteiro S. The Quality of Assessment of Learning (Qual) score: validity evidence for a scoring system aimed at rating short, workplace-based comments on trainee performance. *Teach Learn Med*. 2020 May 26;32(3):319-29. <https://doi.org/10.1080/10401334.2019.1708365>
 45. Warm EJ, Edgar L, Kelleher M, et al. A guidebook for implementing and changing assessment in the milestones era [Internet]. *Accreditation Council for Graduate Medical Education*; 2020, p. 21. Available from: <https://www.acgme.org/Portals/0/Milestones%20Implementation%202020.pdf> [Accessed Oct 2, 2020].
 46. Lang C, Siemens G, Wise A, Gasevic D. *Handbook of Learning Analytics: First Edition*. Society for Learning Analytics Research. 2017. <https://doi.org/10.18608/hla17>
 47. 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):178-87. <https://doi.org/10.1002/aet2.10087>
 48. 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. <https://doi.org/10.24059/olj.v16i3.273>
 49. Bandi V, Shisong W, Thoma B. CBD Dashboard UI [Internet]. GitHub. 2019 Available from: <https://github.com/kiranbandi/cbd-dashboard-ui> [Accessed on Jun 28, 2019].
 50. Watling CJ, Lingard L. Grounded theory in medical education research: AMEE Guide No. 70. *Med Teach*. 2012;34(10):850-61. <https://doi.org/10.3109/0142159X.2012.704439>
 51. Bandi V. CBD Dashboard UI. GitHub. 2019.
 52. Bismil R, Dudek NL, Wood TJ. In-training evaluations: developing an automated screening tool to measure report quality. *Med Educ*. 2014;48(7):724-32.

- <https://doi.org/10.1111/medu.12490>
53. Thoma B, Caretta-Weyer H, Schumacher DJ, et al. Becoming a deliberately developmental organization: using competency-based assessment data for organizational development. *Med Teach*. in-press.
 54. Downing SM. Validity: on the meaningful interpretation of assessment data. *Med Educ*. 2003;37(9):830-7.
<https://doi.org/10.1046/j.1365-2923.2003.01594.x>
 55. Thoma B, Ellaway R, Chan TM. From Utopia through Dystopia: charting a course for learning analytics in competency-based medical education. *Acad Med*. 2021;
 56. Cook DA, Zendejas B, Hamstra SJ, Hatala R, Brydges R. What counts as validity evidence? Examples and prevalence in a systematic review of simulation-based assessment. *Adv Health Sci Educ*. 2014 May 1;19(2):233-50.
<https://doi.org/10.1007/s10459-013-9458-4>
 57. Orr CJ, Sonnadara RR. Coaching by design: exploring a new approach to faculty development in a competency-based medical education curriculum. *Adv Med Educ Pract*. 2019 May 1;10:229-44.
<https://doi.org/10.2147/AMEP.S191470>
 58. Steinert Y. Faculty development: From rubies to oak. *Med Teach*. 2020 Apr 2;42(4):429-35.
<https://doi.org/10.1080/0142159X.2019.1688769>

Appendix A.

Detailed presentation of the thematic analysis including themes, subthemes, categories, and quotation.

| Theme | Subtheme | Category | Quotation |
|--|--------------------------------------|---|---|
| 1. Analyses of assessments (Figure 1-2, Video 1) | | | |
| | 1.1 Quantitative acquisition metrics | | |
| | 1.1.1 | Average EPA score of individual faculty | I think I'm still trying to refine it and the thing that influenced my feedback most was my average entrustment score and realizing how far to the right I was. It actually--since then, I've been a little bit more aware of when it is okay, so to speak, to give the not fives and when it is okay for me to dictate which EPAs we're doing which was maybe a big learning point for the residents. (Faculty member 1) |
| | 1.1.2 | Average EPA score of entire faculty | I mean the average EPA score, that could be really good or really bad. Like that's gonna have to come I think with context for the program. You're getting very junior learners at a junior time and all your average EPAs are 5. That doesn't compute (Faculty development expert 1) |
| | 1.1.3 | Number of EPAs completed | So nice having the total number of EPAs served and then especially if you filter it by day having it for the period because we usually do a check-in every three months with the competence committee to see who's completing EPAs, who maybe needs some prompting for actually getting their EPAs done or for getting more helpful comments so it's good to have that, of that filter. (Program leader 2) |
| | 1.1.4 | Number of EPAs expired | We have some faculty that are notorious for letting EPAs go expired. So I think that's a big thing to focus on for faculty development just to make sure that if they're continuously having expired EPAs, send them an email and say, "Hey I noticed that you've had several EPAs expire for our residents. If you could just remember to complete them on shift with the resident or as close to the shift as possible, that would be extremely helpful because our competency committee depends on this data." (Program leader 2) |
| | 1.1.5 | Distribution of scores | I think I'm gonna be informed by the things that already exist. So knowing the number of EPA's I filled out and my range of scores that I provide relative to the rest of the faculty, gives me a sense of where I'm performing. (Program leader 1) |
| | 1.1.6 | Percentage of EPAs filled out vs sent | So, from an individual side, you obviously have your own individual data. So you would have this nicely from a faculty development side, I think this would be useful to highlight the percentage being high, whether it's a small number or a large number of total EPA's that have been sent or filled out, then percentage is gonna be useful. Just like the way it is. (Program leader 1) |
| | 1.1.7 | Time to EPA completion | One thing I thought would be useful is maybe to like do a thing where the time or number of days or whatever, that you fill out the EPA from when they submitted it. I think it kinda holds you accountable and is super interesting for yourself to know and lately, I think from that session trying to open it up earlier, trying to get on top of that earlier and talk to the residents during shifts, so that was also very useful, too, a good idea. (Faculty member 2) |
| | 1.2 | Narrative acquisition metrics | I would think so because I mean I can certainly see that what the paper is saying, that you know, number of words would have a correlation with the quality of the feedback, but we know people who can speak well and people who can, you know, need more words to say. But ultimately it has to-- for me, ultimately it has to be meaningful to the residents and to the CCC to make their decision, right? So if there was any way to capture residents' opinion on this or the CCC's assessment of that feedback would be really helpful. I would think, going forward (Faculty development expert 1) |
| | 1.2.1 | Words per comment | I really feel that anytime you just kind of give one number, like an average thing, that only gives limited picture, right? So if there was any way to provide some measure of dispersion whether it's standard deviation or range, whatever else you wanted to provide, like this is the average EPA score but this is their whatever it might be. And same thing for the average words. I think that gives a bit more meaningful information. (Faculty development expert 2) |
| | 1.2.2 | Average words per comment | I think the average words per comment are helpful. I'm just seeing that now for the first time. But you can clearly see the difference in comments between people that have a higher word count versus people who have a lower word count. So I think that would be one way to stratify out which faculty we may need to work on. (Program leader 2) |
| | 1.2.3 | Quality of comments | I mean the word counting's one thing, but I guess I wonder how much--'cause you can--how much is put in there 'cause you can write a lot of stuff and say very little, right? So, I guess there's a balance between sorta not saying enough and sort of brevity, I guess, basically, right? So, I wonder if it's not necessarily the number of words versus how much it's the quality of actually what's written, basically, right? I know the thesis or the thought is that if you write more, you're probably giving better quality EPAs to the residents, but I don't know that word count is the absolute be all end all, basically. (Faculty member 3) |
| 2. Contextualization of analyses (Figure 1-5, Video 1) | | | |

| | | |
|-------------|--|---|
| 2.1 By peer | I love the comparisons because I think as physicians we're quite competitive and sometimes it helps us to know where we are in the pack; to know just how much time and effort we'd have to put into something to pull up our socks. But I would want a way to help them, and I think you did show me this, where they can track themselves over time to see how well they're improving on specific parameters. (Faculty development expert 3) | |
| 2.1.1 | Like I feel maybe if it was a graph of all the scores individually that they've ever given out and then you could hover over the comment that was associated with each of the scores. (Program leader 2) | |
| Overview | | |
| Graph | | |
| | 2.1.1.1 | it'd be very easy to identify people who have high proportions of expired EPAs (Program leader 1) |
| | People who have high proportions of expired EPAs | |
| | 2.1.1.2 | it'd be very easy to identify ... people who have very very high and low deviation in their EPA scores... (Program leader 1) |
| | People who have very high and low deviation in their EPA scores | |
| | 2.1.1.3 | it'd be very easy to identify ... people who have very very small word counts. (Program leader 1) |
| | People who have very small word counts | |
| | 2.1.1.4 | Honestly, I think it's interesting 'cause there's so many EPAs, but if it was like you've never done an EPA on this or--'cause if you keep picking the same ones, that's also kind of interesting like for all your residents, you're really good at getting the x,y, z or 90% of residents are missing this and that sort of stays up, but maybe it would have to be location specific, so at the U of S, our residents have the hardest time with these five EPAs. Then, that sort of almost a little just reminder piece and you could scroll over them and maybe read them 'cause I do find it had. I know some general description of the ones from that workshop, but are the four point five point whatever, is that in my brain? No. I don't actually know how to go back and find that. (Faculty member 4) |
| | 2.1.1.5 EPA expiring changes overtime (e.g., Histogram) | So I think that's a big thing to focus on for faculty development just to make sure that if they're continuously having expired EPAs, send them an email and say, "Hey I noticed that you've had several EPAs expire for our residents. If you could just remember to complete them on shift with the resident or as close to the shift as possible, that would be extremely helpful because our competency committee depends on this data." So I think that's the first thing is just having them do the EPA. (Program leader 2) |
| | 2.1.1.6 | Cause obviously there's a few people at the top end who clearly work with residents a lot, and then the rest of them aren't separated by a ton, so it'd be useful for them to know, well, this person far left might have only had one shift, or maybe they've had 50 shifts. And that's dramatically different than the fact that they've only filled out four EPA's. And so from an individual level, from just an accountability side of things, to be like oh boy, I'm at one EPA per four shifts and we could have like a target on that one that said, you know, "The goal of our residency program is a minimum of one EPA per emerge shift, and you as a faculty member are filling out an average of 0.3 per emerge shift. You're below the target". (Program leader 1) |
| | 2.1.1.7 | I think that would be good. The only temporality a graph would offer is if we do do some faculty development it'd be people who were like, five, five, five, five, five initially and we're actually like, no maybe take a minute, think about what comes into the five. Think about what you're doing with the EPA getting some prompting on kind of proper EPA procedures and then maybe there are threes, fours, they have a better mix afterwards. That might be a little bit more valuable but. (Program leader 2) |
| | 2.1.1.8 | So I have to admit, you know again going back to my specialty, 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 and the EPA rating (Faculty development expert 2) |
| | Expired EPA filter | |

| | | | |
|--|-----------------|---|---|
| 2.2 By rotation | 2.1.1.9 | Comparing average EPA score of individual faculty to the overall norm | I like the comparative to see where you are. I think that's very useful just to see where you sit amongst your peers. Like I said, being an outlier, being in the middle doesn't necessarily mean anything, but it stimulates thought and reflection as to where you might wanna put some effort into change. And yeah, so you've got the number you've provided. (Program leader 2) |
| | 2.1.1.10 | Individual faculty to national norm | What we have here is I'm comparing to USask Faculty, but how do I compare it to faculty, nationally, right? And what's sorta the right mix, so are we an anomaly. Are we doing things right or more than right or less than right, so that would be interesting to me, as well, like not only how do I compare locally, but also like nationally, basically, right? At least taking the cohort of emergency physicians that are filling these out from outside of Saskatchewan or outside of our group. (Faculty member 3) |
| | 2.2.1 | Prioritizing EPAs for faculty | And similarly with faculty we could look at that rotation and say, "Look, this is the rotation we really need to get this one observed. Can your faculty please target it?" (Program leader 1) |
| | 2.3 By time | | |
| | 2.3.1 | Date | I think that is helpful. That gives you a little bit more of the background and the ability to compare a certain time frame. It definitely gives you information on how you're doing in terms of EPAs filling out and stuff. (Faculty member 4) |
| | 2.3.2 | Pre-evaluation, post-evaluation trend and comparison | And then almost like putting a marker, a date marker on here of like your filter could be since you received your evaluation, right? On this day. Like pre-evaluation, post-evaluation trend and comparison. (Program leader 1) |
| | 2.3.3 | Words filter (e.g. number of words threshold) | I don't know if there's a way to filter it. Like overall if you could do a global filter of all faculty and then put it like anybody who has less than 30 words per comment and then have those faculty come up and start there with targeted faculty development I think could be an option. ... one way to stratify that would be okay anybody that's done comments with less than 20 words, maybe I could look at those. They're probably all gonna be lesser quality comments. Just based on what I've seen going through the data. And then I could have that list of those people and then I could target them individually. 'Cause I think especially with having off-service faculty on the system, like I think there's gonna be a plan in the future of doing detailed faculty development with our E.R. group and doing this on top of that, targeting off-service people as well I think would be super helpful. And that would be a quick way to filter it. (Program leader 2) |
| | 2.4 By resident | | |
| | 2.4.1 | Resident | Being able to sort by EPA, by resident you worked with, by the date, and by the rating, I think the EPA and the rating are gonna be more useful. 'Cause you can look and see, okay, when did I give ones and twos? When did I give fours and fives? And how often am I doing that? And then specific if there's an EPA that seems to be challenging for me, then I can focus in on that and see where my score ranges are. That would be useful. (Program leader 1) |
| | 2.4.2 | Resident Tier | I think it would be maybe even a breakdown of which tier of residents you evaluate, right? So, core versus end of residency versus early. (Faculty member 1) |
| 2.5 By assessment system | | | So, I'm just imagining something that could kind of show you some semblance of this, but also with the focus of which (EPAs) have you done more and less of relative to everyone else... It just gives you somewhere to focus, I think, too, to help them get that specific learning experience or to really seek it out and be on the lookout for it 'cause there are gonna be rare ones that you could be like, 'Oh, did we see it this shift? So, let's make sure we go take it and take the opportunity to fill out an EPA on it.' (Faculty Member 4) |
| 3. Accessible and clear reporting (Supplemental Data, Video 1) | | | |
| 3.1 Accessible Reporting | | | |
| | 3.1.1 | Exportable Report card | It'd be helpful to have this be downloadable as some sort of a report that could be given to faculty with context or discussed. (Faculty development expert 1) |

| | | |
|---------------------|--|--|
| | | Well I actually think that it would be better to get like a report card or feedback from the competence committee chair or the program director 'cause I think it's kinda like the residents, where you can look at this and it's in a vacuum, right? And it may or may not mean something too. (Faculty development expert 1) |
| 3.2 Clear Reporting | 3.1.2.1 Frequency of report card | I think the ability to login is great, but for me, I am sadly living in a world where I need things pushed into my face, sometimes multiple times, so I think the Rob and Lindsay workshop was super helpful because it was like, "Here's your report. Let's talk about it." So, at least having-- I don't think I need that now because I understand it and plus this session with you, looking at those extra things that have been added, but, now, I would be able to, if I had a report pushed at me, I would be like oh, okay, here's my annual, bi-annual--maybe bi-annual's probably fine. I feel like you want it frequently enough to look at it, but then if it's coming out every month, it becomes something that you delete, you know what I mean? Definitely I think a push method would be better and somewhere between quarterly and annual, but I think monthly would be too much. (Faculty member 4) |
| | 3.1.1 Faculty dashboard login | You could also track yourself in the way of, obviously, things you're doing and how you're comparing up to the group? 'Cause that would--it's your data right? This is probably shared with the Program Directors and probably yourself, basically, right? But, that would--yeah, that might be interesting to look at your--almost like your faculty score or whatever you're calling this, basically. (Faculty member 3) |
| | 3.2.1 Mouse-over Explanations | And then, the other is, is it possible, when I mouse over 4.2, it would tell me what 4.2 is? Because when I look at this, I'm gonna--as a faculty, I'm gonna look at this and say, "Oh geez, I never fill out EPA blah, blah, blah," right? What is that? And then, I could just mouse over the number 3.8 and see, oh okay, that's--3.8, I forget. It's like tox or obstetrical or whatever it is, but yeah okay, I can start filling more of that one out or whatever else, as opposed to having to go to a separate list... Because when I look at this, I'm gonna--as a faculty, I'm gonna look at this and say, "Oh geez, I never fill out EPA blah, blah, blah," right? What is that? And then, I could just mouse over the number 3.8 and see, oh okay, that's--3.8, I forget. It's like tox or obstetrical or whatever it is, but yeah okay, I can start filling more of that one out or whatever else, as opposed to having to go to a separate list." (Program Leader 1) |
| | Legend: EPA: Entrustable Professional Activity, CCC: Competence Committee Chair, CBD: Competence By Design | |

ACQUISTION METRICS FOR ALL FACULTY IN ROTATION - EM

TOTAL EPAS
OBSERVED

2880

PERCENTAGE OF
EPAS EXPIRED

14%

AVERAGE EPA
SCORE

4.16

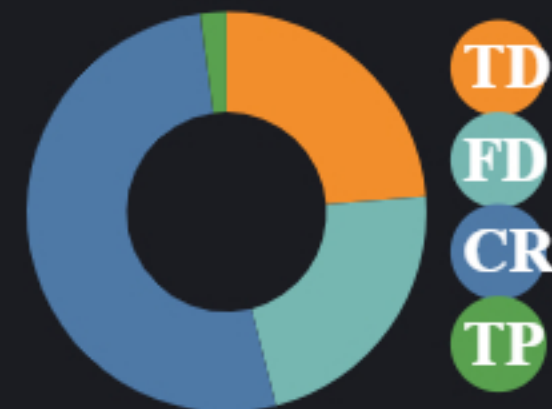
AVERAGE WORDS
PER COMMENT

54

EPA RATING



TRAINING STAGE



ACQUISTION METRICS FOR FACULTY - BRENT THOMA

TOTAL EPAS
OBSERVED

244

PERCENTAGE OF
EPAS EXPIRED

5%

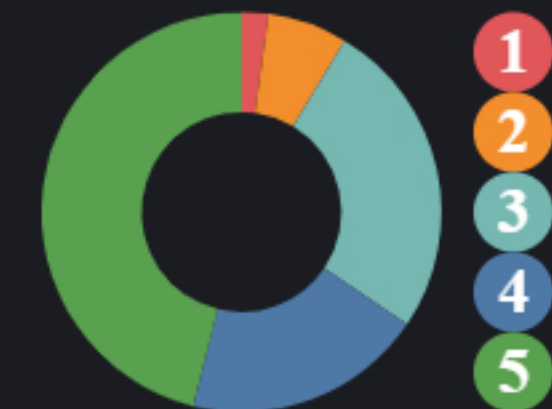
AVERAGE EPA
SCORE

4.01

AVERAGE WORDS
PER COMMENT

88

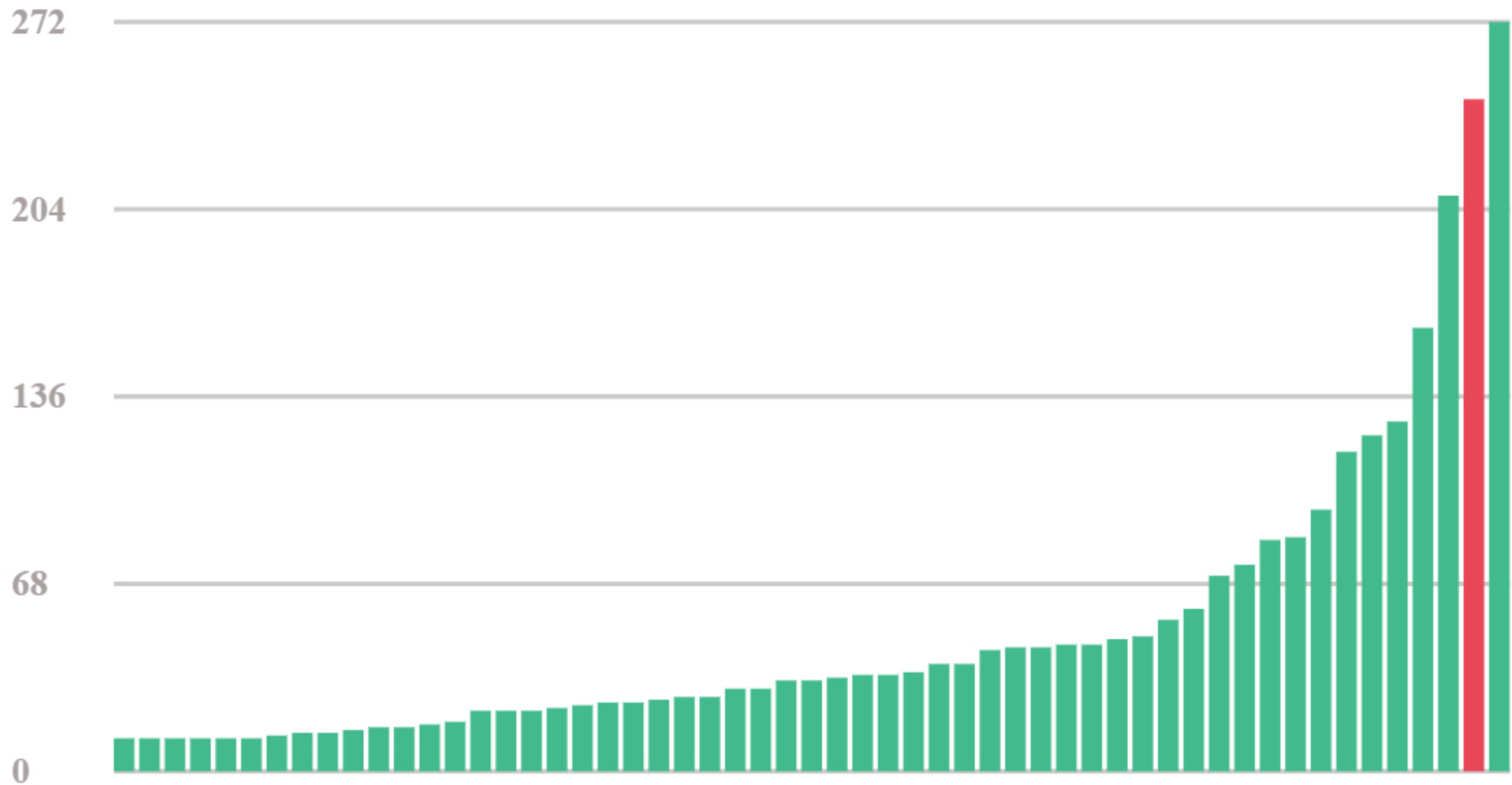
EPA RATING



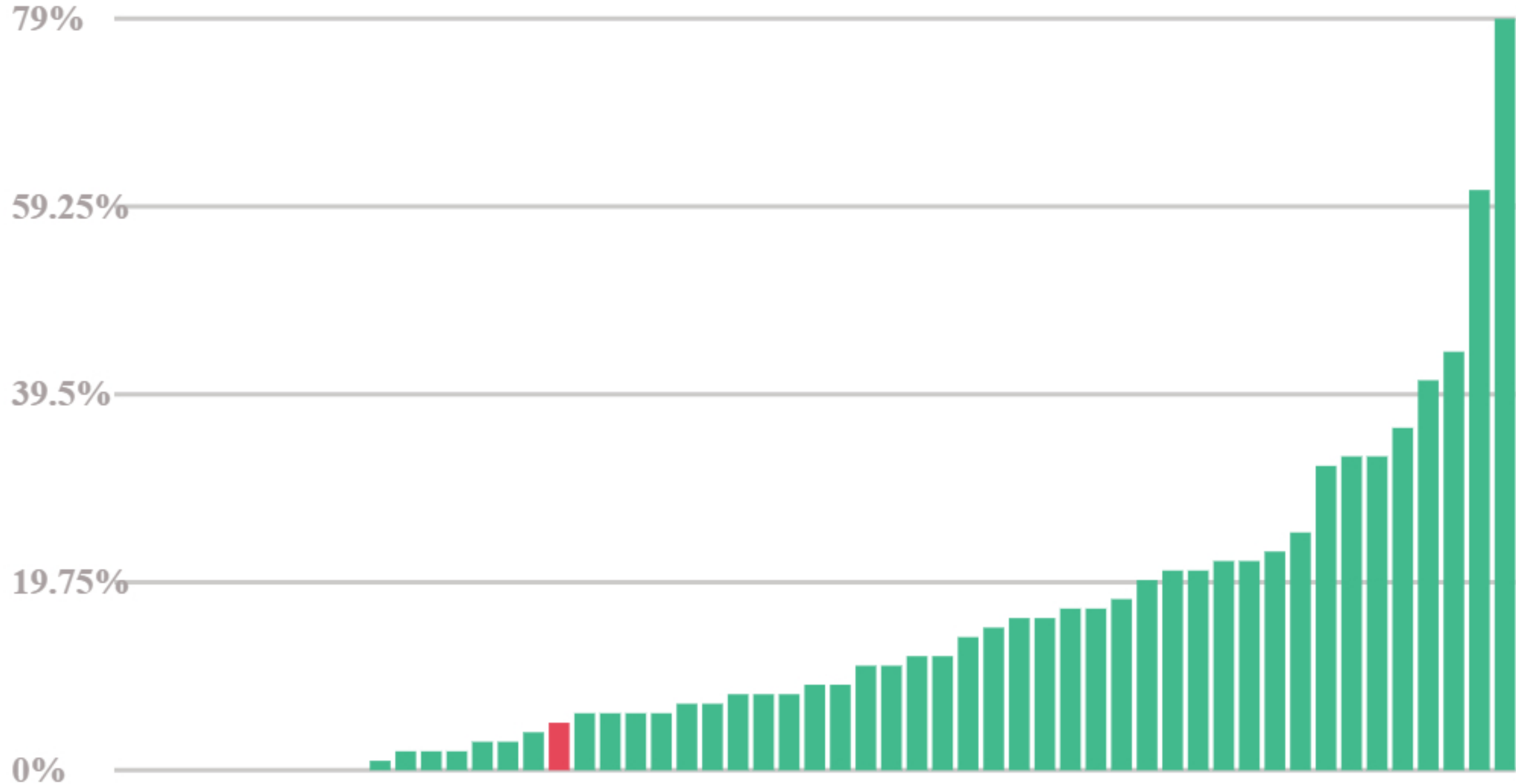
TRAINING STAGE



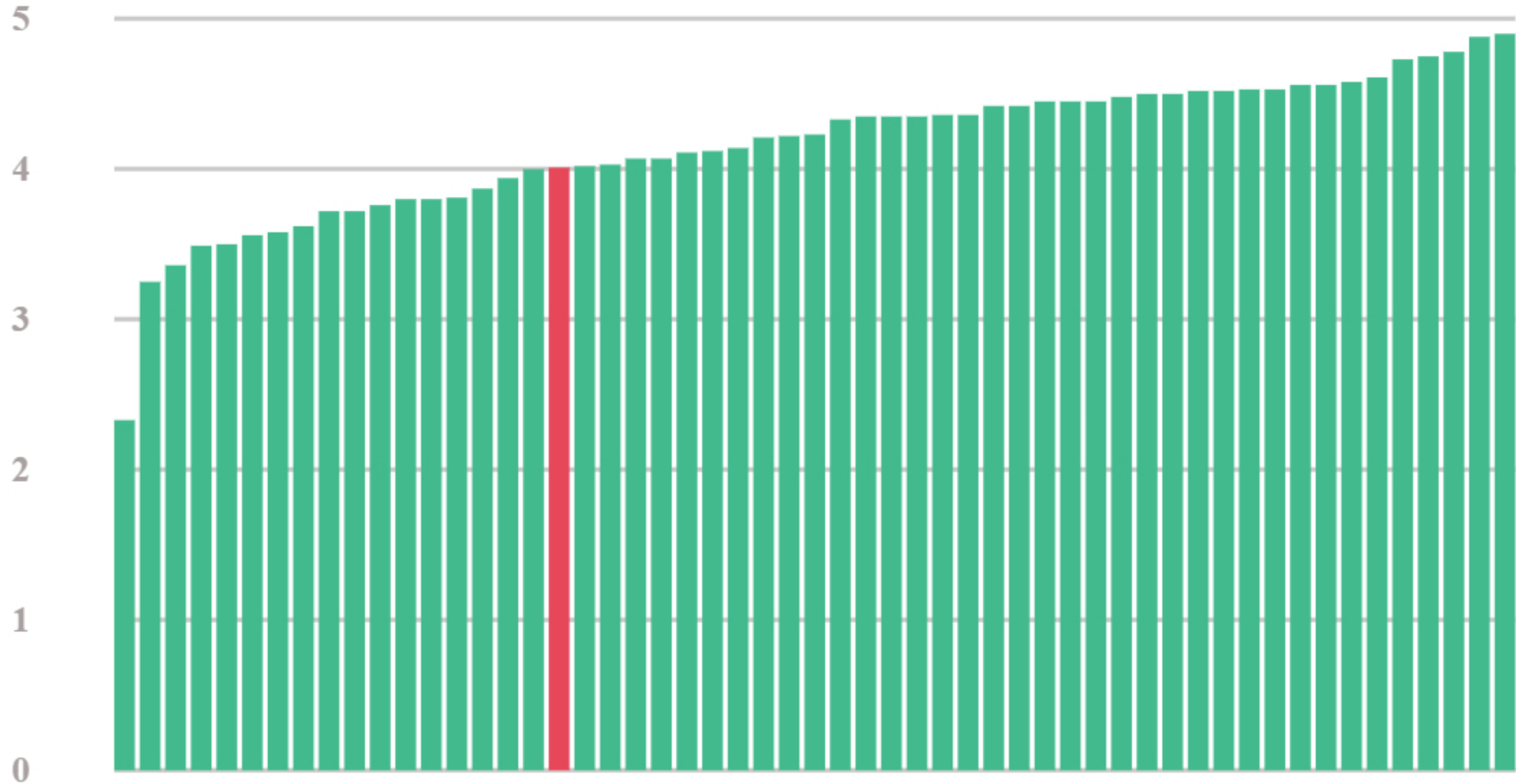
Total EPAs Observed i **244**



EPA Expiry Rate ⓘ 5%



Average Entrustment Score ⓘ **4.01**



Average Words Per Comment i **88**

