

2016

# To submit or not submit: The burden of evaluation on postgraduate medical trainees

Elaine M. Zibrowski

*The University of Western Ontario*

Jeff Crukley

*University of Western Ontario, jcrukley@uwo.ca*

Jacqui Malett

*Western University*

Kathryn Myers

*The University of Western Ontario*

Follow this and additional works at: <https://ir.lib.uwo.ca/fimspub>



Part of the [Library and Information Science Commons](#)

---

## Citation of this paper:

Zibrowski, E., Crukley, J., Malett, J., & Myers, K. (2016). To submit or not submit: The burden of evaluation on postgraduate medical trainees. *MedEdPublish*, 5.

## To submit or not submit: The burden of evaluation on postgraduate medical trainees

Elaine Zibrowski[1], Jeff Crukley[2], Jacqui Malett[3], Kathryn Myers[4]

**Corresponding author:** Dr Kathryn Myers [kathryn.myers@sjhc.london.on.ca](mailto:kathryn.myers@sjhc.london.on.ca)

**Institution:** 1. University of Western Ontario, 2. University of Toronto, 3. University of Western Ontario, 4. University of Western Ontario

**Categories:** Assessment, Research in Medical Education, Technology

Received: 02/08/2016

Published: 05/08/2016

### Abstract

#### Purpose

Academic centers utilize web-based surveillance systems to administer their evaluations, but little is known about their impact on the evaluation responsibilities delegated to medical residents.

#### Method

Using a mixed-methods approach, a retrospective content analysis was conducted of the evaluation activities experienced by a cohort of 29 residents as they completed their training in general internal medicine from 2009-2012. These data were triangulated with group interviews conducted with current internal medicine residents in 2012-2013.

#### Results

The internal medicine program electronically requested that its residents complete 8,614 evaluation reports on clinical faculty, curriculum, and junior trainees (345 requests annually per resident).

Residents reported feeling overwhelmed by their ongoing evaluation workload, and admitted that their motivation to submit high-quality appraisals was dissipating. Residents perceived that their program valued certain evaluations more than others, and this was a major factor in their decision regarding whether or not they would eventually submit an appraisal. Feedback submitted on program evaluation-related appraisals were viewed as having the least value, and residents were significantly less likely to submit these evaluations.

#### Conclusions

Although web-based surveillance systems are efficient in distributing thousands of evaluations, residency programs to engage in ongoing vigilance of the unintended consequences associated with their use.

**Keywords:** postgraduate, rater fatigue, assessment, technology

### Article

## Introduction

Many North American academic centres, including our own, introduced web-based surveillance networks to administer its evaluations of clinical faculty, trainees, and medical education programs (Civetta et al 2001; Deretchin et al 2002; D’Cunha et al 2003; Bennett et al 2004; Feldman & Triola 2004; Natt et al 2006). While it was anticipated that the shift from a hard-copy dependent to electronic platform would reap benefits including greater compliance, improved data quality, and increased access to evaluative information, to date, the literature has been focused on implementation processes, rather than outcomes, associated with these surveillance systems.

We argue that meaningful assessment systems can evolve only through careful exploration of their facets and complexities including the workflow, and impact of these activities, experienced by its user groups. Undoubtedly, one user group involves medical residents, who are delegated the professional responsibility to serve as evaluators throughout the tenure of their postgraduate education. While we have previously reported that residents struggle to meet their obligations as evaluators (Myers et al 2012), we know little about their actual evaluation workloads. Moreover, while ‘survey fatigue’ has been thought to negatively impact survey return rates (Porter et al 2004), the degree to which medical residents’ experience and react to these feelings, within the context of an electronic surveillance system, has not been previously explored. The purpose of this study was to explore the evaluation workflow and habits experienced by postgraduate trainees during residency training.

## Methods

We conducted a mixed-methods study involving a content analysis of the evaluation activities undertaken by a historical cohort of 29 general internal medicine (GIM) residents. From 2009-2012, these residents completed a three-year residency program at a single academic centre in London, Canada. The unit of analysis was the individual, electronically delivered evaluation form that the residents received throughout each academic year by means of the electronic software program used by the department’s Education Office. Using this web-based software, GIM residents’ received email prompts alerting them to the arrival of an evaluation form to their individual electronic account along with a timed reminder that an appraisal was due. Evaluation requests sent by other administrative units to the members of the historical cohort, including those from external rotations, and requests sent from the dean of postgraduate education, were not included in the sample. Text-based records relating to academic year, evaluation form name, and whether or not it submitted was by the resident were exported into an Excel spreadsheet in August, 2012. Submitted evaluation forms were not made available to the research team, and residents’ data was individually aggregated via generic code. Blank copies of the forms were used to record its characteristics including type of evaluation and its number of items. The formatted dataset was then imported to SPSS for analyses (IBM SPSS Statistics for Mac, Version 22). We supplemented the secondary evaluation data with information gathered from semi-structured, group interviews with GIM trainees from August 2012 - January 2013. Although the residents who participated in the interviews were not members of the historical cohort, they were representative members of the very next group of trainees to undergo GIM training, and therefore, were similarly exposed to evaluation practices as per their immediate predecessors.

Residents were sent email inviting them to consider participation in the study if they had received at least one evaluation request to from the Education Office’s software program. Group interviews were

scheduled by convenience, and were conducted by a trained researcher who did not know the interviewees. The interview guide was informed by the by initial analyses of the secondary dataset, and by our previous research in rater-based assessment (Myers et al 2012; Zibrowski et al 2011; Myers et al 2011). We audio-recorded the interviews, and these were transcribed into verbatim electronic documents. Interviewees received a \$25 gift card as an honorarium. Over the course of interviews we used an inductive, constant comparative approach so that the data was reviewed after each session, by a minimum of two members of the research team. This allowed the interview guide to be edited to maximize capture of ideas emerging within and across the group discussions.

We summarized secondary evaluation records with mean and standard deviation and/or percent response, and minimum and maximum values. 95% confidence intervals were calculated around point estimates. In order to explore the potential relationship between evaluation form type and year in residency program (PGY), we used a chi-square test of association and the relationship between the total number of items probed by an evaluation and whether or not the form was submitted was estimated by a point-biserial correlation with 95% confidence interval.

## Results

The Education Office sent a grand total of 8,614 evaluation requests to the 29 residents from July 1, 2009 to June 30, 2012. On average, resident received, 344  $\pm$ 126 forms [minimum = 148, maximum = 578; 95% CI (341.3, 346.7)]. These involved three major types of requests: a) *program evaluation (academic session evaluation, OSCE evaluation, rotation evaluation)*: 5529/8614 (64.2 %); b) *residents' appraisal of a physician acting in a supervisory role (clinical teacher, consultant, senior resident)*: 3006/8614 (34.9%); and c) *residents' appraisal of a junior trainee (mid-rotation clerkship evaluation, physical exam evaluation, medical dictation evaluation)*: 79/8614 (0.9%). The mean number of items on the set of forms was 18 (7 SD, minimum = 4, maximum = 45). A statistically-significant relationship was detected between postgraduate year of training and type of evaluation request with more requests sent out during year one ( $\chi^2_4=172.8$ ;  $p<.001$ ). The majority of these were for appraisals of clinical faculty and academic sessions (83%). Seven residents, one PGY1, three PGY2, three PGY3, were interviewed in the group sessions. All spoke frustratingly about the plethora and volume of evaluation requests they continually received. *"They can really accumulate and then you can feel overwhelmed by the different ones."* (R1) *"These tend to be tremendously onerous to the point where it's completely impractical to do them. Because you just do not have that much time"*. (R2) A couple of the residents complained about the electronic system not being user-friendly, and it seemed to complicate their access to the forms, *"I could do it on my phone right away, like, if they just send me the form on my phone. But I have to go log in on a computer."* (R6) *"It's a horrible website"*. (R5) Moreover, some trainees commented on how the evaluation tasks competed for attention with other aspects of their life. As this resident rationalized, *"The reason I don't do them is that we get caught up in everyday things. You come from work, you're tired, you have a couple errands to do and then you know, you find the dinner if you can, and you're tired. You just want to go to bed"*. (R7)

Residents submitted a grand total of 7,428 (86.2%) of the forms they received. Individual submission totals varied from 100% (8/29 residents) to 3.4% (1/29 residents). When asked about their motivation to submit evaluations, the trainees who were interviewed admitted that their motivation to submit evaluations had declined since entering residency, and that this was impacted, to some degree, by the volume of requests: *"I mean having gone through these for a few years, I would spend less time on them than I used to"*. (R3). *"I definitely let it go over a year or more. I knew I had them there. I didn't really*

*forget them. I did keep putting it off and by nature I'm bit of a procrastinator anyways. I ended up getting to the point just where it looked pretty overwhelming". (R4)*

That being said, some residents acknowledged that their program's evaluation system seemed to be set up to monitor and encourage residents' completion of certain forms, such as ratings of clinical faculty. "You will not get your evaluation if you do not complete theirs. So, those tend to get done a little more quickly for me". (R1) "Our program director says, 'you're not filling out your online evaluation form' and you're thinking, alright, alright, I guess, I'll go home and sort that out." (R5) A weak correlation was detected between number of items probed and submission of an evaluation [ $r_{pb} = -.12$ ;  $p < .001$ ; 95% CI (-.14 to -.10)].

Residents' openly described their personal strategies for managing the volume of forms they received including: "I do them right away. You just want to get it over with". (R2) "I think the highest number I got to was 115 in my box. Eventually I did do these over an afternoon slash evening. I think I did most of them or entered all of them at once ". (R4)

In their discussion about their personal strategies, residents admitted that even though an evaluation was technically submitted back to the GIM program, its quality might not be optimal. All of the residents who were interviewed described shortcutting strategies they used to complete their evaluations as quickly as possible, "I can honestly say I've never elected to evaluate more than one person" (R3), "I'm just getting it done. Just getting the thing done and even if it's horrible, even if it's not great, it's quick and I answer down the middle ". (R6)

A significant relationship was also found between type of evaluation form and whether or not it was submitted ( $\chi^2 = 427.6.5$ ;  $p < .001$ ). Ninety percent of the forms that were not submitted were program evaluation-related, with the majority of these being requests for an appraisal of an academic session (994/1078). When asked about how they decide whether to submit or not submit an evaluation, residents revealed that a major factor in these decisions appeared to be the value they anticipated their program would have for the feedback. As this resident indicated, "My perception is that for faculty evaluation, there is more weight placed on that because it affects how those people are probably ultimately remunerated or promoted. Those things do get fed on and that I think, it does actually get implemented in some way down the road in terms of how people perform on service. But I think, when you're talking about the didactic lectures that are given there's been a tremendous amount of feedback over the years of how about these need to be improved and really there's been very minimal change". (R3) Moreover, some trainees felt that the web-based program was not the ideal venue for communicating constructive feedback because, "I think there are better venues for that issue to arise rather than on some paper, some electronic document" (R5), and "If I really want to change something then I would go speak to the person that I needed to". (R6).

## Discussion

This study is the first to explore, using naturalistic data, the evaluation workflow of a cohort of general internal medicine residents. Based on our examination of these data, we theorize that the expectations placed on residents as evaluators is undermining their ability to complete high-quality reports, and perhaps more importantly, is at odds with the overarching goal of the evaluation system, that is, continuous quality improvement of the residency program.

We found that the evaluation software enabled the GIM training program to place considerable workloads onto its trainees, as they were sent, on average, nearly 345 forms to complete per year. It should be noted that this estimate took only into account internal requests; it did not include those made

annually by other administrative units and programs. Therefore, at minimum, they were asked to contemplate nearly 19,000 questions during their residency experience. While it could be argued that trainees are not required to answer every question posed to them on an evaluation, and therefore, a given form should take only a few minutes to complete, less than one-third of the historical cohort submitted all of their forms. Moreover, the residents interviewed by this study, expressed that the volume of their evaluation workload left them feeling overwhelmed and frustrated. Rather than feeling an enhanced sense of compliance, they sensed their overall motivation to continually work on these forms was dissipating. They admitted to using shortcuts, which we have described previously (Myers et al 2012), to help just ‘*get it done*’, and in some cases the form was not submitted at all. Porter and colleagues (2004) commented that even in situations where the number of surveys individuals are asked to complete is low, the process of sending multiple, overlapping requests can stimulate feelings of response burden or ‘survey fatigue’. Although we were unable to aggregate the evaluation requests sent per month/rotation across the historical cohort, the total number sent to these trainees, their tendency to submit fewer forms as their training progressed, along with remarks from interviewees’ regarding accumulations in their e-accounts, suggests that residents commonly encounter situations where multiple forms are competing for their attention. As such, residency programs need to be vigilant to thresholds for requests made to residents, beyond which impacts the reliability and validity of the appraisals they make. Future research into the mental workload expended by postgraduate trainees as raters is warranted.

Griffin and Cook (2009) remarked that many higher education programs have problems ‘*closing the feedback loop*’, with their stakeholders, and often do not make explicit the link between feedback and institutional change. They posit that in order for stakeholders to remain engaged, and trust in the evaluation process, they must experience some outcomes of their feedback. The residents’ reactions to, and the submission frequencies for, certain evaluations, like the ratings of academic half-days, raises the question of how well the program is responding to its data. We have previously reported that residents’ lack knowledge on the evaluation process, and sense that their training program valued some appraisals, like clinical teaching evaluations, more than others (Myers et al 2012). The results of this study further suggest that these perceptions influence their decisions about whether or not they will ultimately submit an evaluation. We have advocated that residency programs invest in educating its trainees on how it collects, stores, and disseminates evaluative data to its clinical faculty and higher administration (Myers et al 2012). We build on this, and suggest that efforts made to enhance transparency also need to include delivery of tangible evidence to postgraduate trainees of how their residency program will prioritize and enact meaningful improvements, in response to their critiques.

Although web-based software can be very effective in distributing a surfeit of evaluation requests, the results of our study highlight the need for residency programs to engage in ongoing vigilance of the unintended consequences associated with their use. Although we conducted this study within a single academic center, the prevalence of electronic evaluation systems within medical education suggests that our results will likely resonate with other schools, but further research could enhance the transferability of our findings. Future insight may also be gained through exploration at the level of the residency program to explore how they manage, review and evaluate the data in order to close the evaluative feedback loop.

## **Take Home Messages**

---

- Residents' evaluation workload can lead them to become feeling overwhelmed and frustrated.
- Departmental expectations competes with residents' ability to continually complete high-quality, evaluation reports.
- Residency programs should engage in ongoing vigilance of the unintended consequences associated with their use of web-based, evaluation systems.

## Notes On Contributors

---

Elaine Zibrowski is a population epidemiologist, and a research associate in the Department of Medicine, Schulich School of Medicine & Dentistry. She is a PhD candidate in Health Information Science, University of Western Ontario, London, Canada.

Jacqui Malett is a resident in general internal medicine, Schulich School of Medicine & Dentistry, University of Western Ontario, London, Canada.

Jeff Crukley is an audiologist, and an adjunct professor, in speech-language pathology in the Faculty of Medicine, University of Toronto, Toronto, Canada.

Kathryn Myers is a professor of medicine, and chair/chief of General Internal Medicine, Schulich School of Medicine & Dentistry, University of Western Ontario, London, Canada.

## Acknowledgements

---

The authors express gratitude to the Department of Medicine's Education Office, and to the residents who were interviewed during this study. We also thank Tavis Apramian, Holly Ellinor, and Joan Windler for their assistance. This study was funded by a *Faculty Support for Research in Education Grant* from the Schulich School of Medicine & Dentistry. During this study's data collection and preliminary analysis phases, E. Zibrowski was the Manager & Research Consultant and J. Crukley was a Research Associate at the Centre for Education Research & Innovation (CERI), Schulich School of Medicine & Dentistry.

Ethical approval: This study received approval from the Health Sciences Research Ethics Board (Delegated Review) at the University of Western Ontario.

Previous presentations: Abstracts based on earlier versions of this paper were presented at the 2012 CERI Annual Research Symposium; 2013 Department of Medicine Resident Research Day; 2013 Canadian Conference in Medical Education (CCME); and the 2013 International Conference on Residency Education (ICRE).

## Bibliography/References

---

Bennett, AJ, & Arnold, LM. Use of a computerized evaluation system in a psychiatry clerkship. *Acad Psych*. 2004; 28(3): 197-203.

<http://dx.doi.org/10.1176/appi.ap.28.3.197>

Civetta, JM, Morejón, OV, Kirton, OC, Reilly, PJ, Serebriakov, II, Dobkin, ED & Antonetti, M. Beyond requirements: residency management through the Internet. *Arch Surg*. 2001; 136(4): 412-417.

<http://dx.doi.org/10.1001/archsurg.136.4.412>

D'Cunha, J, Larson, CE, Maddaus, MA, & Landis, GH. An internet-based evaluation system for a surgical residency program. *J Amer Coll Surg*. 2003; 196(6): 905-910.

[http://dx.doi.org/10.1016/S1072-7515\(03\)00111-X](http://dx.doi.org/10.1016/S1072-7515(03)00111-X)

Deretchin LF, Wheeler DA, and Jefferson, LS. A Web-based evaluation system. *Acad Med* 1997; 72:418-9.

<http://dx.doi.org/10.1097/00001888-199705000-00042>

Feldman, HJ, & Triola, MM. ResEval: a web-based evaluation system for internal medicine house staff. In Proceedings of the 13th international World Wide Web conference on Alternate track papers & posters. 2004; May: 336-337.

<http://dx.doi.org/10.1145/1013367.1013463>

Fortin, AH, Luzzi, K, Galaty, L, Wong, JG, & Huot, SJ Developing an Internetbased Communication System for Residency Training Programs. *JGIM*. 2002; 17(4): 278-282.

<http://dx.doi.org/10.1046/j.1525-1497.2002.10737.x>

Griffin, A. & Cook, V. Acting on evaluation: Twelve tips from a national conference on student evaluations. *Med Teach*. 2009; 31(2): 101-104.

<http://dx.doi.org/10.1080/01421590802225788>

Myers, KA, Zibrowski, EM, & Lingard, L. A mixed-methods analysis of residents' written comments regarding their clinical supervisors. *Acad Med*. 2011; 86(10): S21-S24.

<http://dx.doi.org/10.1097/ACM.0b013e31822a6fd3>

Myers, K, Zibrowski, EM, & Lingard, L. Engaged at the extremes: Residents' perspectives on clinical teaching assessment. *Acad Med*. 2012; 87(10): 1397-1400.

<http://dx.doi.org/10.1097/ACM.0b013e3182674488>

Natt, N, Dupras, DM, Schultz, HJ, & Mandrekar, JN. Impact of electronic faculty evaluation on resident return rates and faculty teaching performance. *Med Teach*. 2006; 28(2): e43-e48.

<http://dx.doi.org/10.1080/01421590600603608>

Porter, SR, Whitcomb, ME, & Weitzer, WH. Multiple surveys of students and survey fatigue. *New Direct Instit Res*. 2004; 121: 63-73.

<http://dx.doi.org/10.1002/ir.101>

Zibrowski, EM, Myers, K, Norman, G, & Goldszmidt, MA. Relying on others reliability: challenges in clinical teaching assessment. *Teach Learn Med* 2011; 23 (1): 21-27.

<http://dx.doi.org/10.1080/10401334.2011.536754>

## Appendices

---

## Declaration of Interest

---

*The author has declared that there are no conflicts of interest.*