

Feedback Credibility in a Formative Postgraduate Objective Structured Clinical Examination: Effects of Examiner Type

Lynfa Stroud, MD, MEd

Matthew Sibbald, MD, PhD

Denyse Richardson, MD, MEd

Heather McDonald-Blumer, MD, MSc

Rodrigo B. Cavalcanti, MD, MSc

ABSTRACT

Background Resident perspectives on feedback are key determinants of its acceptance and effectiveness, and provider credibility is a critical element in perspective formation. It is unclear what factors influence a resident's judgment of feedback credibility.

Objective We examined how residents perceive the credibility of feedback providers during a formative objective structured clinical examination (OSCE) in 2 ways: (1) ratings of faculty examiners compared with standardized patient (SP) examiners, and (2) ratings of faculty examiners based on alignment of expertise and station content.

Methods During a formative OSCE, internal medicine residents were randomized to receive immediate feedback from either faculty examiners or SP examiners on communication stations, and at least 1 specialty congruent and either 1 specialty incongruent or general internist faculty examiner for clinical stations. Residents rated perceived credibility of feedback providers on a 7-point scale. Results were analyzed with proportional odds models for ordinal credibility ratings.

Results A total of 192 of 203 residents (95%), 72 faculty, and 10 SPs participated. For communication stations, odds of high credibility ratings were significantly lower for SP than for faculty examiners (odds ratio [OR] = 0.28, $P < .001$). For clinical stations, credibility odds were lower for specialty incongruent faculty (OR = 0.19, $P < .001$) and female faculty (OR = 0.45, $P < .001$).

Conclusions Faculty examiners were perceived as being more credible than SP examiners, despite standardizing feedback delivery. Specialty incongruency with station content and female sex were associated with lower credibility ratings for faculty examiners.

Introduction

Formative assessment is an essential component of competency-based medical education in graduate medical education.¹⁻³ Parallel to this is the increasing appreciation for the constructivist nature of the feedback process.⁴⁻⁶ For example, individual perceptions of the feedback,⁷ its alignment with self-assessment,⁸ and trainee opinions of feedback providers^{7,9} are important filters through which information is viewed and used.^{7,10-12} Perceived credibility of feedback providers is a major determinant of how recipients view and incorporate feedback.^{9,13-17} Higher perceived credibility is associated with a greater likelihood that recipients view feedback as accurate,¹⁸ useful,¹⁹ and actionable.¹⁶ Credibility may be enhanced by feedback provider characteristics, such as seniority, clinical expertise, and trustworthiness.¹³ Conversely, feedback providers with poor interpersonal skills or lack of attention to processes may be viewed as less credible.¹³

Credibility of feedback providers is highly relevant in competency-based medical education. Both the Accreditation Council for Graduate Medical Education and the Royal College of Physicians and Surgeons of Canada mandate the incorporation of nonphysician feedback on some domains of assessment.^{20,21} Within simulated assessments, such as formative objective structured clinical examinations (OSCEs), this may include feedback from standardized patients (SPs). It is important to consider how learners perceive nonphysician providers of feedback, and how they judge their credibility.

Formative OSCEs offer beneficial educational opportunities that trainees value.^{16,22} Concurrently, OSCEs are characterized by limited interaction between feedback recipient and provider. Most studies of determinants and effects of credibility have focused on longitudinal feedback recipient-provider relationships.⁵ It is possible that different patterns of credibility perceptions exist in an OSCE setting, due to a heterogeneity of providers and shorter interactions, compared with clinical environments.

DOI: <http://dx.doi.org/10.4300/JGME-D-17-00578.1>

Exploratory data collected during a previous OSCE suggested that residents perceived faculty examiners as being more credible than SP examiners, and viewed general internist faculty as being more credible than subspecialists.²³ However, confounding factors may have influenced these findings. We designed this study to systematically examine factors that influence residents' perceptions of the credibility of feedback providers on a formative OSCE.

Methods

Setting

We conducted this study in 2015, during an annual formative OSCE for internal medicine residents at the University of Toronto. We aimed to compare feedback experiences from a variety of providers: generalist faculty, subspecialist faculty, and SPs.

Residents rotated through 2 identical tracks: 2 structured cases to assess clinical skills; a physical examination station with an SP; and a communication scenario with an SP. Faculty examiners (general internists and medical subspecialists) provided feedback on clinical and physical stations, and typically SPs provided feedback on communication stations without faculty involvement. Global scores were assigned using a 5-point scale anchored by level of training (1, clerk, to 5, consultant); residents were blinded to scores and received only comments. Residents received 4 minutes of feedback immediately after each station. Raters provided feedback that was specific, action-oriented, and short (maximum 3 items). Raters watched a training video and participated in an orientation prior to the OSCE.

Intervention

We blueprinted the examination for clinical and physical stations. Each resident received feedback from faculty examiners whose expertise was congruent with station content (specialty congruent faculty) and from specialty incongruent faculty or general internists. Faculty explicitly introduced themselves to residents as a generalist or a specialist, including specialty type. For communication stations, residents were randomized to receive feedback from either an SP or a faculty examiner. Faculty observed the interaction behind a 1-way mirror to control who generated the feedback, and to standardize feedback delivery. As the SP exited the room, their feedback was conveyed to faculty, who then presented this feedback to residents as their own. To confirm accuracy, SPs observed feedback provision from behind the 1-way mirror. In the SP feedback track, SPs exited the room and returned to provide feedback.

What was known and gap

Acceptability and actionableness of feedback to trainees is critical in competency-based medical education.

What is new

Internal medicine residents in a formative OSCE rated the faculty credibility for clinical content, and standardized patient feedback on communication skills.

Limitations

Single institution, single specialty study limits generalizability.

Bottom line

Faculty examiners were perceived as being more credible than SP examiners, and specialty congruent examiners were rated as being more credible, with specialty incongruency examiner female sex associated with lower credibility ratings.

Immediately following the examination, residents completed a questionnaire. For each station, residents documented if their examiner was a generalist, specialist (type), or SP. They rated the credibility of the feedback provider on a 7-point Likert scale (1, poor, to 7, excellent). A 7-point scale also was used to mitigate ceiling effects. For faculty examiners, residents indicated previous contact in the clinical setting. We also collected demographic variables about feedback providers and residents.

The study was approved by the University of Toronto Research Ethics Board.

Statistical Analysis

We characterized participants using descriptive statistics, and we examined the characteristics of each dyadic resident-rater interaction for both clinical/physical and communication stations. We used univariate linear regression to assess relationships between perceived credibility of examiner, resident sex, and level of training. We also evaluated whether resident performance related to his or her ratings of credibility by comparing univariate analyses for residents receiving low (< 3 of 5) versus high (≥ 3 of 5) scores on any station. Through univariate analyses we explored whether raters' attributes were related to perceived credibility. We constructed a proportional odds model using credibility ratings as the dependent variable, with the ratings categorized into 3 levels: ≤ 5 , 6, and 7 (including all collected variables as predictors). To ensure repeated measures were accounted for in the clinical station model, we built a clustered proportional odds regression model that accounts for clustering through a random intercept for each resident. Proportional odds models were chosen due to concerns regarding ceiling effects in credibility scores leading to nonnormal distributions. We included examiner familiarity into the model, because prior research suggested it has

TABLE 1

Examiner Demographics on Dyadic Interactions With Residents During Postgraduate Formative Objective Structured Clinical Examination

Examiners	Dyadic Interaction With Resident, n (%)
Sex	
Male	375 (49)
Female	393 (51)
Type (all stations)	
SP	107 (14)
Faculty	661 (86)
Type (communication station)	
SP	107 (56)
Faculty	85 (44)
Faculty congruency ^a	
Generalist	100 (17)
Congruent	252 (44)
Incongruent	224 (39)
Prior clinical familiarity ^a	
Yes	90 (16)
No	479 (83)

Abbreviation: SP, standardized patient.

^a Faculty-only stations.

important effects in OSCE scoring.^{2,3} For communication stations, we constructed a similar multivariable proportional odds model. All analyses were performed using R version 3.4.0 (The R Foundation for Statistical Computing, Vienna, Austria).

Results

Descriptives

We analyzed data for 192 of 203 residents (95%), 72 faculty, and 10 SPs. Residents were balanced by level of training (74 PGY-1, 58 PGY-2, and 60 PGY-3) and sex (41% female [79 of 192]). Data were available for 768 dyadic resident-examiner interactions (TABLE 1).

Univariate Analysis

We examined the relationship between resident and examiner demographic variables, and resident perceptions of examiner credibility (TABLE 2). There was a small but significant tendency for male residents to rate the credibility of faculty examiners lower than female residents (6.2 versus 6.4, $P = .030$), and for credibility ratings to be lower as residents progressed through training (PGY-1 = 6.4; PGY-2 = 6.3; PGY-3 = 6.2; $P = .002$). We found no significant differences in residents' credibility judgments of SPs based on resident sex and PGY level. There were also no significant differences in the perceived credibility of faculty examiners for residents who received low

TABLE 2

Univariate Analyses of Association of Resident and Examiner Characteristics With Perceived Credibility of Faculty Feedback Providers

Characteristic	Credibility Mean (IQR)	P Value
Residents		
Sex		.030
Male	6.2 (6–7)	
Female	6.4 (6–7)	
Level		.002
PGY-1	6.4 (6–7)	
PGY-2	6.3 (6–7)	
PGY-3	6.2 (6–7)	
Performance		.75
≥ 3	6.3 (6–7)	
< 3	6.3 (6–7)	
Examiners		
Sex		< .001
Male	6.4 (6–7)	
Female	6.2 (6–7)	
Examiner type		< .001
Standardized patient	5.4 (4–6)	
Faculty	6.3 (6–7)	
Faculty congruity		< .001
Congruent	6.6 (6–7)	
Incongruent	6.0 (5–7)	
Generalist	6.5 (6–7)	
Familiarity		.35
Yes	6.4 (6–7)	
No	6.3 (6–7)	

Abbreviations: IQR, interquartile range; PGY, postgraduate year.

($n = 208$) compared with high ($n = 549$) station scores.

Residents perceived the credibility of feedback providers differently based on examiner demographics. Male faculty were rated as being more credible than female faculty (6.4 versus 6.2, $P < .001$), but there were no significant differences between male and female SPs. There was no sex interaction effect ($P = .37$). There was a significant difference in credibility perceptions for faculty examiners' congruency with station (congruent = 6.6; generalist = 6.5; incongruent = 6.0; $P < .001$), but there was no effect of familiarity with a faculty examiner on perceptions of credibility (see the FIGURE for credibility ratings of faculty by sex and specialty congruence). At the communication stations, faculty examiners were perceived as being more credible than SP examiners (6.3 versus 5.4, $P < .001$). In a restricted analysis of

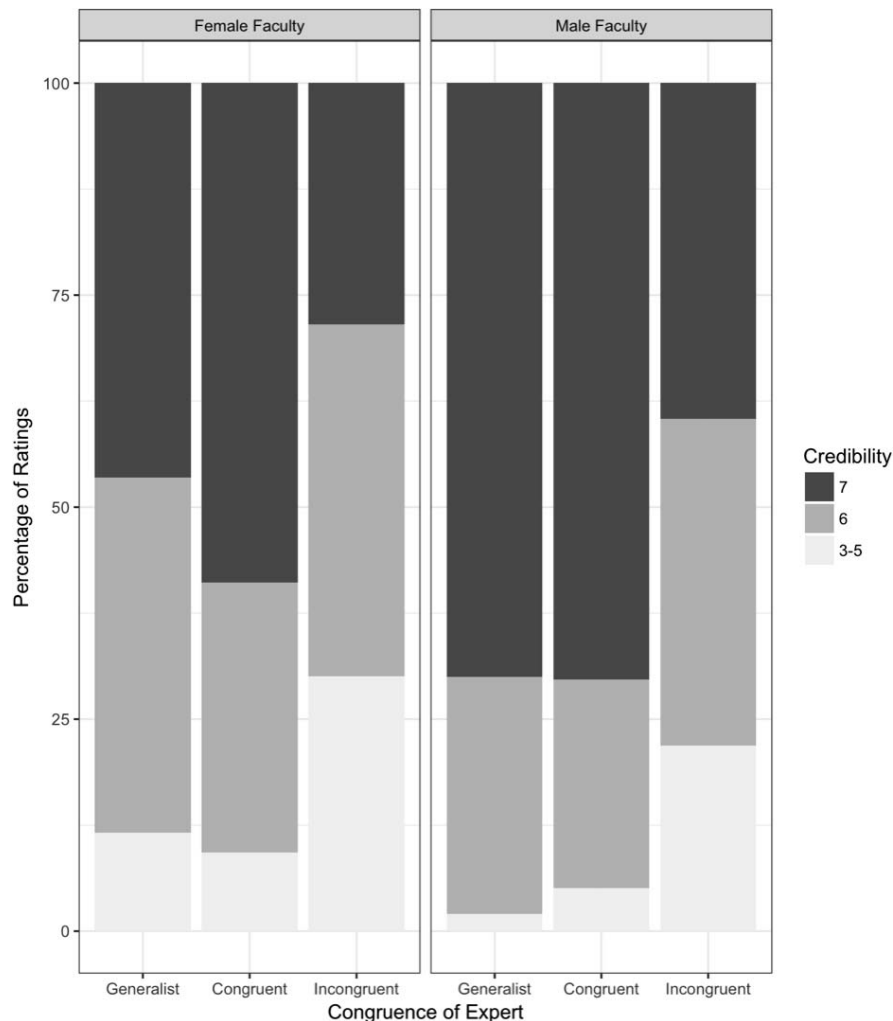


FIGURE
Credibility Ratings by Sex and Specialty Congruency on Clinical Stations

examiners who received low credibility ratings (≤ 4 of 7), similar patterns were replicated.

Multivariable Analysis

We constructed 2 multivariable proportional odds regression models (1 for faculty-only stations and 1 for communication stations) to assess the effects of demographic variables on perceived examiner credibility (TABLES 3 and 4). Results were similar to those from univariate analyses. For faculty-only stations, specialist incongruent faculty had significantly lower odds of high credibility, with an odds ratio (OR) of 0.19 (95% confidence interval 0.10–0.37). This suggests general internists had 5 times higher odds of receiving high credibility ratings compared with examiners whose specialty did not align with station content. In contrast, no significant differences were found for specialty congruent raters compared with generalists. Male sex was associated with higher odds

of high credibility ratings (OR = 2.21, $P = .002$; for female faculty, OR = 0.45, $P < .001$). On the communication stations, only examiner type was significantly associated with perceived credibility, with SP raters having reduced odds of receiving a higher credibility rating (OR = 0.28, $P < .001$).

Discussion

Residents rated the credibility of all examiners highly in this formative OSCE, even when differences existed (the lowest mean rating was 5.4 of 7). This is reassuring, as credibility of feedback providers is associated with greater acceptance and utilization of feedback.¹⁶ However, differences in residents' credibility judgments of feedback providers existed, with the greatest differences observed between faculty and SP examiners on communication stations. As faculty credibility ratings were similar on clinical (mean = 6.3 of 7; median = 7; interquartile range

TABLE 3

Odds Ratios for Receiving a Higher Credibility Rating on Proportional Odds Model by Characteristics Associated With Perceived Credibility of Faculty Feedback on Clinical Stations

Variable	Odds Ratio for Higher Credibility Rating (95% CI)	P Value for Term	P Value for Whole Variable
Resident level (c/w PGY-1)			.14
PGY-2	0.80 (0.33–1.97)	.63	
PGY-3	0.42 (0.17–1.02)	.06	
Resident sex (c/w female)			
Male	0.54 (0.25–1.15)	.11	
Examiner sex (c/w female)			
Male	2.21 (1.35–3.61)	.002	
Examiner congruity (c/w generalist)			< .001
Specialist congruent	1.74 (0.91–3.32)	.09	
Specialist incongruent	0.19 (0.10–0.37)	< .001	
Familiarity			
Yes	1.67 (0.85–3.30)	.13	

Abbreviations: CI, confidence interval; c/w, compared with; PGY, postgraduate year.

[IQR] = 6–7) and communication (mean = 6.3 of 7; median = 6; IQR = 6–7) stations, this suggests that differences are not domain related. Rather, it indicates differences in how residents perceive the credibility of physician versus nonphysician feedback providers.

Mechanisms driving resident perceptions are not entirely clear. However, the theories of self-categorization and social identity may offer an explanation.²⁴ These frameworks suggest that residents “identify” with faculty examiners as their “in-group,” and categorize SPs as an “out-group.” This in turn may influence perceptions of each group as feedback providers.

While social identity theory may partly explain our findings, other factors likely underpin credibility judgments. Residents may view faculty and SP examiners’ credibility differently based on perceptions of what the OSCE represents.²⁵ Ultimately, residents will be examined by physicians on their

certification examination. Thus, they may view formative OSCEs as preparation for this, placing greater value on faculty examiners’ feedback. Additionally, credibility judgments may not remain stable over time. While univariate analyses suggested later stages of training may lead to lower credibility ratings, in our multivariable proportional odds model accounting for repeated measures, this was not statistically significant. This association may warrant further investigation.

Finally, perceived credibility of faculty examiners was influenced by their congruity with station content, with the highest credibility ratings given for station-congruent faculty, followed by general internists, followed by station-incongruent faculty. Perceived expertise is a well-documented key element of credibility,^{13,17} and it is not surprising that faculty examining at a station within their area of specialty were rated as most credible. However,

TABLE 4

Odds Ratios for Receiving Higher Credibility Rating on Proportional Odds Model by Characteristics Associated With Perceived Credibility of Faculty Feedback on Communication Stations

Variable	Odds Ratio for Higher Credibility Rating (95% CI)	P Value for Term	P Value for Whole Variable
Resident level (c/w PGY-1)			.30
PGY-2	0.77 (0.40–1.47)	.43	
PGY-3	0.60 (0.31–1.15)	.12	
Resident sex (c/w female)			
Male	0.79 (0.46–1.37)	.41	
SP sex (c/w female)			
Male	1.07 (0.63–1.84)	.80	
Examiner type (c/w faculty)			
SP	0.28 (0.16–0.48)	< .0001	

Abbreviations: CI, confidence interval; c/w, compared with; PGY, postgraduate year; SP, standardized patient.

generalist internists were rated almost as highly. This may be due to their perceived broad level of expertise, or due to residents identifying with generalists, as they were in an internal medicine training program. That station-incongruent faculty examiners were rated as being significantly less credible than either group may have implications for examiner/station blueprinting.

There are limitations to our study. The number of dyadic interactions with SPs were fewer than with faculty, which may have been insufficient to reach significance for some variables in these analyses. Similarly, we may have been unable to determine the effects of familiarity on credibility for faculty examiners. We did not include a measure of residents' perceptions about faculty seniority and its influence on credibility, previously documented to be important.¹³ We did not attempt to measure residents' perceptions of faculty communication skills. The residents may perceive faculty to be better communicators than SPs, and this may have influenced our findings.

We also did not explore drivers of residents' different perceptions of credibility, or how perceptions of credibility affected acceptance or utilization of feedback. As we move toward incorporating the input of nonphysicians into formative assessments with the implementation of competency-based medical education, determining how residents perceive this feedback will be critical to maximizing its effectiveness and mitigating any potential unintended consequences.

Conclusion

Faculty examiners were perceived as being more credible than SP examiners, despite standardizing feedback delivery. Specialty incongruity with station content and female sex resulted in lower credibility ratings, with sex having a small but real effect. Gaining a better understanding of residents' perceptions of credibility, in particular with respect to nonphysicians, and exploring what drives these perceptions is an important next step.

References

- Holmboe ES, Sherbino J, Long DM, et al. The role of assessment in competency-based medical education. *Med Teach*. 2010;32(8):676–682.
- The Future of Medical Education in Canada. A collective vision for postgraduate medical education in Canada. 2012. https://afmc.ca/future-of-medical-education-in-canada/postgraduate-project/pdf/FMEC_PG_Final-Report_EN.pdf. Accessed February 12, 2018.
- Busing N, Harris K, MacLellan A, et al. The future of postgraduate medical education in Canada. *Acad Med*. 2015;90(9):1258–1263.
- Rushton A. Formative assessment: a key to keep learning? *Med Teach*. 2005;27(6):509–513.
- Telio S, Ajjawi R, Regehr G. The “educational alliance” as a framework for reconceptualising feedback in medical education. *Acad Med*. 2015;90(5):609–614.
- Stone D, Heen S. *Thanks for the Feedback: The Science and Art of Receiving Feedback Well*. New York, NY: Penguin Books; 2014.
- Watling CJ, Lingard L. Toward meaningful evaluation of medical trainees: the influence of participants' perceptions of the process. *Adv Health Sci Educ Theory Pract*. 2012;17(2):183–194.
- Sargeant J, Mann K, van der Vleuten C, et al. “Directed” self-assessment: practice and feedback within a social context. *J Contin Educ Health Prof*. 2008;28(1):47–54.
- Eva KW, Armson H, Holmboe E, et al. Factors influencing responsiveness to feedback: on the interplay between fear, confidence, and reasoning processes. *Adv Health Sci Educ Theory Pract*. 2012;17(1):15–26.
- Archer JC. State of the science in health professional education: effective feedback. *Med Educ*. 2010;44(1):101–108.
- Sargeant JM, Mann KV, van der Vleuten CP, et al. Reflection: a link between receiving and using assessment feedback. *Adv Health Sci Educ Theory Pract*. 2009;14(3):399–410.
- Eva KW, Munoz J, Hanson MD, et al. Which factors, personal or external, most influence students' generation of learning goals? *Acad Med*. 2010;85(suppl 10):102–105.
- Bing-You RG, Paterson J, Levine MA. Feedback falling on deaf ears: residents' receptivity to feedback tempered by sender credibility. *Med Teach*. 1997;19(1):40–44.
- Watling CJ, Kenyon CF, Zibrowski EM, et al. Rules of engagement: residents' perceptions of the in-training evaluation process. *Acad Med*. 2008;83(suppl 10):97–100.
- Sargeant J, Mann K, Ferrier S. Exploring family physicians' reactions to multisource feedback: perceptions of credibility and usefulness. *Med Educ*. 2005;39(5):497–504.
- Sargeant J, Eva KW, Armson H, et al. Features of assessment learners use to make informed self-assessments of clinical performance. *Med Educ*. 2011;45(6):636–647.
- Watling C, Driessen E, van der Vleuten CPM, et al. Learning from clinical work: the roles of learning cues and credibility judgments. *Med Educ*. 2012;46(2):192–200.
- Albright MD, Levy PE. The effects of source credibility and performance rating discrepancy on reactions to

multiple raters. *J Appl Soc Psychol*. 1995;25(7):577–600.

19. van de Ridder JMM, Berk FCJ, Stokking KM, et al. Feedback providers' credibility impacts students' satisfaction with feedback and delayed performance. *Med Teach*. 2015;37(8):767–774.
20. Accreditation Council for Graduate Medical Education. Common Program Requirements. <http://www.acgme.org/What-We-Do/Accreditation/Common-Program-Requirements>. Accessed February 12, 2018.
21. Royal College of Physicians and Surgeons of Canada. General standards applicable to all residency programs. B standards. <http://www.cmq.org/pdf/agreement/normes-b-agreements-ang-2013.pdf>. Accessed February 12, 2018.
22. Losh D. The value of real-time clinician feedback during OSCEs. *Med Teach*. 2013;35(12):1055.
23. Stroud L, Herold J, Tomlinson G, et al. Who you know or what you know? Effect of examiner familiarity with residents on OSCE scores. *Acad Med*. 2011;86(suppl 10):8–11.
24. Burford B. Group processes in medical education: learning from social identity theory. *Med Educ*. 2012;46(2):143–152.
25. Harrison CJ, Konings KD, Dannefer EF, et al. Factors influencing students' receptivity to formative feedback

emerging from different assessment cultures. *Perspect Med Educ*. 2016;5(5):276–284.



Lynfa Stroud, MD, MEd, is Associate Professor, Department of Medicine, and a Centre Researcher, Wilson Centre, University of Toronto, Toronto, Ontario, Canada; **Matthew Sibbald, MD, PhD**, is Assistant Professor, Department of Medicine, and Director, Centre for Simulation-Based Learning, McMaster University, Hamilton, Ontario, Canada; **Denyse Richardson, MD, MEd**, is Associate Professor, Department of Medicine, University of Toronto; **Heather McDonald-Blumer, MD, MSc**, is Associate Professor, Department of Medicine, University of Toronto; and **Rodrigo B. Cavalcanti, MD, MSc**, is Associate Professor, Department of Medicine, and Director, HoPingKong Centre for Excellence in Education and Practice, University Health Network, University of Toronto.

Funding: This study was funded by an education scholarship and research grant from the Department of Medicine, University of Toronto.

Conflict of interest: The authors declare they have no competing interests.

The authors would like to thank George Tomlinson, PhD, for his expert statistical assistance and advice, and Lisa St Amant, research assistant, for her help with this project.

Corresponding author: Lynfa Stroud, MD, MEd, Sunnybrook Health Sciences Centre, D4-70A, 2075 Bayview Avenue, Toronto, ON M4N 3M5, Canada, 416.480.6100 ext 83627, lynfa.stroud@sunnybrook.ca

Received August 11, 2017; revision received October 31, 2017; accepted December 11, 2017.