

Multisource Feedback Tool for the Assessment of Medical Student Clerks in Professionalism, Communication, and Collaboration Skills

Ahmed Al Ansari^{1,2} · Aysha Waheed Agab³ · Sarah Dawood Al Sayed⁴ · Noora Jassim Al Fudhala⁵

Published online: 6 September 2016

© International Association of Medical Science Educators 2016

Abstract

Construct This study investigates the validity of the Bahrain Defence Force instrument to assess professionalism, communication, and collaboration skills (BDF/PCC instrument) in medical graduates during their clerkship/intern years.

Approach The instrument to assess professionalism, communication, and collaboration skills [BDF/PCC instrument] was developed based on an extensive literature review, other existing valid instruments, and expert opinion. Exploratory and confirmatory factor analysis was conducted to support the construct validity. Twenty-five interns engaged in a multisource feedback evaluation for this study. Each participant was rated by eight individuals from each of the following categories: physicians,

nurses, and fellow interns. Cronbach's α was used to determine the questionnaire's internal consistency and reliability.

Results We report response rates (100 %), mean response time to complete each questionnaire (3.7 min), and the number of raters (seven to eight) needed to provide reliable results that support the feasibility of the survey. Reliability analysis using Cronbach's α of internal consistency indicated that the full scale of the instrument had high internal consistency (Cronbach's α 0.98). The instrument was found to be suitable for factor analysis (KMO = 0.941; Bartlett test significant, $p < 0.001$), which found that the data collected from the questionnaire could be grouped into three factors. These three factors represented 77.3 % of the total variance: professionalism, collaboration, and communication. The item-total correlation for this instrument was above 0.40 and showed homogeneity within each composite scale. The generalizability coefficients (Ep^2) were 0.79 for the surveys.

Conclusion The BDF/PCC instrument to assess professionalism, communication, and collaboration skills is a feasible, reliable, and valid tool to assess physicians in their clerkship year. Testing the instrument on two occasions with a 1-year interval and the confirmatory factor analysis provided some evidence to support the concurrent and the construct validity of the BDF/PCC instrument.

✉ Ahmed Al Ansari
ahmed.alansari@bdfmedical.org; drahmedalansari@gmail.com

Aysha Waheed Agab
a.w.agab@gmail.com

Sarah Dawood Al Sayed
Sarahdawood08@gmail.com

Noora Jassim Al Fudhala
Njf080114@gmail.com

¹ Training and Education Department, Bahrain Defence Force Hospital, P. O. Box - 28743, West Riffa, Kingdom of Bahrain

² Assistant Professor of General Surgery Arabian Gulf University, Manama, Kingdom of Bahrain, Senior Clinical Lecturers in Medical education, RCSI Bahrain, Adliya, Kingdom of Bahrain

³ Ophthalmology Department, Bahrain Defence Force Hospital, Riffa, Bahrain

⁴ Pediatric Department, Bahrain Defence Force Hospital, Riffa, Bahrain

⁵ General Surgery Department, Bahrain Defence Force Hospital, Riffa, Bahrain

Keywords MSF · Reliability · Validity · Clerkship · PCC

Introduction

The internship year can be referred as the formal training period of medical graduates or a transformation period from being a student in class room to being a clinician who cares for patients. In today's healthcare system, it is necessary to have excellent communication, professionalism, and collaboration

skills. Lack of communication creates situations where medical errors can occur and these errors may have the potential to cause severe injury or unexpected patient death. Collaboration between physicians, nurses, and other health care professionals increases the awareness of the type of knowledge and skills of each other member and it helps in leading to continued improvement in decision-making. Professionalism is central to sustaining the public's trust in the medical profession; it is the essence of the doctor–patient relationship. Professionalism is identified as a competency of resident education.

Multisource feedback (MSF) is a popular tool of assessment that relies on the evaluations of different groups of people, often including both physicians and non-physicians. MSF has been established as both a feasible and reliable process to measure physicians' performance on a variety of factors at varying stages of their medical careers [1, 2].

There has yet to be an established standard for the number of assessors that are ideal for the most reliable results; however, it has been found that this is influenced by a number of factors, including the number of items on the questionnaire, the competencies being assessed, and the background of the assessors [3]. For example, an increased number of raters may be needed for assessing interpersonal skills and professionalism, as these are subjective domains that would benefit from input from multiple sources. Multisource feedback is useful for assessing several domains, including professionalism, communication, and collaboration. Because these values and behaviors go beyond the technical knowledge and skills that are usually evaluated in an academic setting, MSF is an essential form of feedback that may be used to improve intern performance [4].

MSF has been proven to be useful in facilitating stronger teamwork [5], increasing productivity [6], and enhancing communication and trust among employees [7]. As such, it is particularly useful in identifying those individuals who have relatively weaker interpersonal skills, allowing for a learning opportunity for intervention to improve the future physician, and it helps them to serve the patients better in future. [4] Studies have shown that, particularly in weaker-scoring individuals, MSF ratings tend to be lower relative to self-ratings and as such may provide eye-opening feedback to those being rated [8].

It previously has been established that when individuals receive feedback on their MSF ratings, they tend to build on the constructive criticism to improve their personal skills and ultimately enhance the medical care that patients receive [9]. MSF is both a reliable, feasible, and time-efficient tool to evaluate medical interns [10], and as such, it was used in this study to assess our group of interns as they ventured out into the field.

Different tools can be used in the MSF process [11]. The quality of the tool in the form of the psychometric properties is important because, later on, organizations will base their decision on the results obtained by using those tools; therefore, such instruments should be feasible, reliable, valid, and have

an education impact. The primary aim of the study therefore was to construct a new tool to assess professionalism, communication, and collaboration and to explore the feasibility, reliability, and validity of the new tool, which is called the BDF/PCC instrument to assess professionalism, communication, and collaboration. The secondary aim of this study was to assess the feasibility of implementing the MSF process in assessing interns in their clerkship year.

Methods

This study was conducted in the Bahrain Defence Force Hospital, a military teaching Hospital in the Kingdom of Bahrain between March 2014 and June 2015. The hospital has 450 beds, 322 physicians and dentists, and 1072 nurses and practical nurses [12]. We conducted the MSF on all 25 interns with a 100 % response rate. There were 17 females and 8 males in the sample.

Instrument Development Process

The BDF/PCC was developed based on extensive literature review and other existing validated instruments such as the physician achievement review instrument (PAR) [13, 14], Maastricht history-taking and advice scoring list (MAAS-Global) instrument, [15] Calgary–Cambridge tool to assess communication skills [16], Sheffield peer review assessment tool (SPRAT) [17], assessment of interprofessional team collaboration scale (AITCS) [18], and experts' opinion in the field of medical education. The focus of the instrument is to assess professionalism, communication skills, and collaboration.

To establish relevance and content validity, a table of specifications based on the literature search and the previous instruments was constructed, and a working group was involved in developing the instrument. Expert opinion was taken into consideration by sending the instrument to five experts, published in the field of medical education and PCC, and asked to judge whether the content and the format of the instrument were appropriate in comparison with the table of specification, which was sent to them.

The instrument consisted of 39 items: 15 items to assess professionalism, 13 items to assess communication skills, and 11 items to assess collaboration skills. The instrument was constructed in a way that can be applied by different groups of people including interns, senior medical colleagues, consultants, and co-workers. The items on the instrument had a five-point Likert response scale in the form of the following: 1 = among the worst; 2 = bottom half; 3 = average; 4 = top half; and 5 = among the best with an option of “unable to assess” (UA). After the committee developed the questionnaires, we sent them to the experts for feedback and items were modified based on this feedback.

For the survey based evaluation to assess professionalism, communication skills and collaboration, the lower cut-off score for the rated domains for this study was based on 25th percentile. On the other hand, the higher cut-off score for the rated domains for this study was based on 75th percentile.

Statistical Analysis

A number of statistical analyses were undertaken to address the research questions posed. Response rates, time required to fill out the questionnaire, and the number of raters required to produce reliable results were used to determine feasibility for the BDF/PCC instrument [14, 17].

For each item on the survey, the percentage of the “unable-to-assess” answers, along with the mean and standard deviation, was computed to determine the viability of the items and the score profiles. Items in which the “unable-to-assess” answer exceeded 20 % on a survey might be in need of revision or deletion, according to previously conducted research [14].

Since this instrument was used on two occasions with a 1-year interval, for group one we used exploratory factor analysis to determine which items on each survey were suitable to group together (i.e., become a factor or scale). Using individual-physician data as the unit of analysis for the survey, the items were inter-correlated using the Pearson product moment correlations. The correlation matrix was then broken down into its main components, and these components were then rotated to the normalized varimax criterion. Each item was assigned to the factor on which they were loaded with a loading factor of at least 0.40. If an item was loaded on more than one factor (cross-loading), the item was assigned to the highest factor where it was loaded. The number of factors to be extracted was based on the Kaiser rule (i.e., eigenvalues >1.0) [19].

The factors or scales established through exploratory factor analysis were used to establish the key domains for improvement (e.g., professionalism), whereas the items within each factor provided more precise information about specific behaviors (e.g., maintains confidentiality of patients, recognizes boundaries when dealing with other physicians, and shows professional and ethical behavior). Physician improvement could be guided by the scores on factors or items.

This analysis made it possible to determine whether the instrument items were aligned with the appropriate constructs (factors) as intended. Instrument reliability (consistency) was assessed. In the second implementation, we used a confirmatory factor analysis to ensure that the factors extracted were similar to those that were extracted during the first implementation. This will support the construct validity of the instrument.

To examine the homogeneity of each composite scale, item-total correlations corrected for overlap were calculated [20]. We consider an item-total correlation coefficient of <0.3 as evidence that the item is not measuring the same construct

measured by other composite scale items. [21] In addition, Pearson’s correlation coefficient was used to estimate the inter-scale correlations, which will determine the degree of overlap between the scales [22].

Internal consistency/reliability was examined by calculating Cronbach’s coefficient for each of the scales and for each factor separately. Cronbach’s coefficient is widely used to evaluate the overall internal consistency for each instrument as well as for the individual factors within the instruments [20]. This analysis was followed by a generalizability analysis to determine the Ep^2 and to ensure there were ample numbers of questions and evaluators to provide accurate and stable results for each intern on each instrument. Normally, an $Ep^2 = 0.70$ suggests that data are stable. If the Ep^2 is below 0.7, it suggests that more raters or more items are required to enhance stability [22]. We further conducted a D study where we estimated the Ep^2 for 1–10 raters.

Raters

Three groups of raters were defined as physicians, nurses, and fellow interns. Participants identified 8 physicians, 8 nurses, and 8 fellow interns as potential raters. The basic criterion was that the participant intern must know and have worked with these potential raters for a minimum of 2 months. Investigators selected 4 nominees from each list such that each clerkship intern was rated by 4 physicians, 4 nurses and 4 peers.

Ethical Approval

The research was approved by the research ethics committee in the BDF hospital. Written consent was obtained from the interns, and verbal consent was obtained from raters. The study was conducted between March 2014 and June 2015.

Results

Instrument

The response rate for our MSF questionnaire was 100 %, including all 25 interns in the program; this cohort was comprised of 17 female and 8 male interns. The mean response time to complete each questionnaire was 3.7 min, and 7 to 8 raters were needed to provide reliable results that illustrate the feasibility of the survey. The participants responded to nearly all items on the questionnaire for this study. However, there were 4 questions (Q27, 28, 36, and 38) that exceeded 20 % of the response “unable to assess” by the raters in the first implementation of the survey in March 2014. After revising the 4 questions, no questions exceeded 20 % of the response “unable to assess.”

Table 1 Descriptive statistics, item analysis, and confirmatory factor analysis for the second group cohort 2

Q		N	M	SD	%UA	Self	SD	Correlated item-total correlation ^a	Factors identified by factor analysis		
									Comm	Colla	Profe
Professionalism											
Q1	Maintains confidentiality of patients.	291	4.25	0.86	5.20	4.66	0.48	0.76			0.76
Q2	Recognizes boundaries when dealing with other physicians	301	4.22	0.90	3.30	4.95	0.45	0.73			0.80
Q3	Recognizes boundaries when dealing with other health care professionals	299	4.20	0.88	3.70	4.90	0.55	0.72			0.80
Q4	Shows professional and ethical behavior	299	4.25	0.90	3.70	4.60	1.10	0.62			0.64
Q5	Is punctual, and performs tasks in a time-appropriate manner	295	4.21	0.94	4.40	4.57	0.51	0.64			0.61
Q6	Is able to handle situations in a professional manner and exhibits self-control, avoiding emotional outbursts in stressful situations	281	3.99	0.86	7.10	4.42	0.81	0.63			0.50
Q7	Respects patient's autonomy and right to be involved in his/her own management	279	4.22	0.83	7.50	4.38	1.16	0.61			0.62
Q8	Is reliable and responsible when performing his/her duties	299	4.23	0.87	3.70	4.85	0.36	0.63			0.69
Q9	Is honest, and handles his/her duties in a dignified manner	294	4.28	0.88	4.60	4.85	0.35	0.58			0.66
Q10	Accepts constructive criticism and develops goals for improvement	267	4.00	0.88	9.80	4.66	0.79	0.63			0.50
Q11	Respects cultural, individual and role differences including age, gender, race, religion, disability, language, sexual orientation, and socioeconomic status.	291	4.28	0.83	5.20	4.80	0.51	0.74			0.57
Q12	Follows institutional policies and procedures	297	4.21	0.84	4.00	4.90	0.31	0.56			0.61
Q13	Arrives on time to scheduled appointments and hospital activities	288	4.22	0.96	5.80	4.57	0.59	0.63			0.72
Q14	Manages healthcare resources efficiently	271	4.15	0.79	9.10	4.61	0.58	0.56			0.60
Q15	Leads with respect and fair treatment of colleagues	289	4.22	0.82	5.60	4.85	0.47	0.61			0.71
Communication											
Q16	Communicates efficiently and in a clear, understandable fashion with colleagues within his/her team.	300	4.20	0.87	3.50	4.80	0.51	0.61			0.56
Q17	Communicates efficiently and in a clear, understandable, and compassionate way with patients.	284	4.21	0.83	6.60	4.71	1.10	0.62			0.58
Q18	Allows the patient to elaborate about his condition	275	4.16	0.85	8.30	4.85	0.35	0.61			0.61
Q19	Communicates efficiently and in a clear, understandable, and compassionate way with patient's families	269	4.14	0.83	9.40	4.71	0.46	0.59			0.71
Q20	Communicates clearly and effectively with other healthcare workers, e.g., nurses	289	4.18	0.84	5.60	4.71	0.46	0.52			0.60
Q21	Explains what is being done for the patient during examination or procedures	264	4.11	0.84	10.4	4.71	0.56	0.53			0.77
Q22	Communicates purpose and results of investigations to patients well.	266	4.06	0.85	10.0	4.71	0.56	0.59			0.81
Q23	Follows up appropriately and in a timely manner on patients' hospital course.	281	4.08	0.91	7.10	4.47	1.16	0.58			0.74
Q24	Communicates management options to patients in a clear, understandable way, taking into account the patients' opinion	261	4.10	0.86	11.0	4.09	0.92	0.61			0.81
Q25	Displays empathy in dealing with patients by eye contact and verbal responses	261	4.13	0.82	11.0	4.60	1.16	0.66			0.70
Q26	Summarizes the information given for the patient in small quantities, with concrete explanations, and understandable language	271	4.04	0.85	9.10	4.42	1.20	0.59			0.67
Q27	Maintains calm in emergency situations, in order to communicate information clearly to his/her seniors.	235	4.00	0.83	16.0	4.38	1.10	0.66			0.59
Q28	Communicates accurate patient information to physicians from other departments when required to do so.	260	4.08	0.84	11.2	4.47	1.16	0.67			0.63
Collaboration											
Q29	Manages to work well as part of a healthcare team	294	4.14	0.92	4.60	4.65	0.58	0.60			0.60
Q30	Facilitates the learning of medical colleagues and co-workers.	268	4.08	0.89	9.60	4.66	0.65	0.54			0.60
Q31	Collaborates well with nurses and other healthcare workers.	295	4.18	0.87	4.40	4.76	0.54	0.55			0.73
Q32	Concerned about the safety of patients and co-workers.	278	4.21	0.82	7.70	4.76	0.53	0.56			0.72

Table 1 (continued)

Q	Professionalism	N	M	SD	%UA	Self	SD	Correlated item-total correlation ^a	Factors identified by factor analysis		
									Comm	Colla	Profe
Q33	Coordinates patient care efficiently	284	4.15	0.86	6.60	4.76	0.43	0.62			0.70
Q34	Collaborates with other healthcare workers in order to achieve optimal patient care	283	4.09	0.88	6.70	4.47	1.16	0.63			0.68
Q35	Participates in a system of call in order to provide care for patients	283	4.14	0.85	6.70	4.52	1.16	0.62			0.62
Q36	Provides appropriate guidance and help to team members on regular bases	270	4.06	0.90	9.20	4.38	1.20	0.60			0.66
Q37	Takes on extra work, when appropriate, to help the team	275	4.02	0.97	8.30	4.85	0.35	0.55			0.59
Q38	Enables the team to achieve agreements for team process and collaborative completion of assignment.	264	4.03	0.91	10.4	4.38	1.20	0.61			0.62
Q39	Participates fully in collaborative process and fulfilled team agreements	274	4.08	0.90	8.50	4.85	0.45	0.69			0.61

N number, M mean, SD standard deviation, UA unable to assess, Comm communication skills, Colla collaboration, Profe professionalism

^a Item-total correlation >0.4 considered as each item is correlated with total score of the instrument

The instrument was found to be excellent for factor analysis (KMO = 0.941; Bartlett test significant, $p < 0.001$), which found that the data collected from the questionnaire could be grouped into three factors. These three factors represented 77.3 % of the total variance: professionalism, collaboration, and communication. The item-total correlation for this instrument was above 0.40 and showed homogeneity within each composite scale (Table 1).

Reliability analysis using Cronbach’s α reliability of internal consistency indicated that the full scale of the instrument had high internal consistency (Cronbach’s α 0.98). The reliability for the factors (subscales) within the questionnaire also had high internal consistency and reliability (Cronbach’s $\alpha \geq 0.96$). G study analysis was conducted employing a single-facet, nested design. The generalizability coefficients (Ep^2) were 0.79 for the surveys. A previous D study we had conducted estimated the EP^2 for 1–10 raters, and we found that for one assessor, $EP^2 = 0.30$, 7 assessors =0.75, 8 assessors =0.78, and for 10 assessors =0.81. The item-total correlations were all above 0.40 for all items within their respective scales (Table 1).

We performed a separate reliability analysis for each raters (a small size of $n = 4$). We calculated the instrument internal consistency reliability (Cronbach’s alpha) (Table 2). The results show the instrument consistently reflects the construct it is measuring.

Table 2 Internal consistency reliability analysis

	Cronbach’s alpha
All raters (doctors($n = 8$), nurses($n = 8$), and fellow students($n = 8$))	0.98
Doctors ($n = 4$)	0.993
Nurses ($n = 4$)	0.959
Fellow students ($n = 4$)	0.952

Cronbach’s alpha >0.7 indicates satisfactory internal consistency reliability

Participants

Twenty-five participants were assessed by raters of different categories for their professionalism, communication, and collaboration, respectively. The interns’ mean scores on professionalism, communication and collaboration are plotted in Fig. 1a.

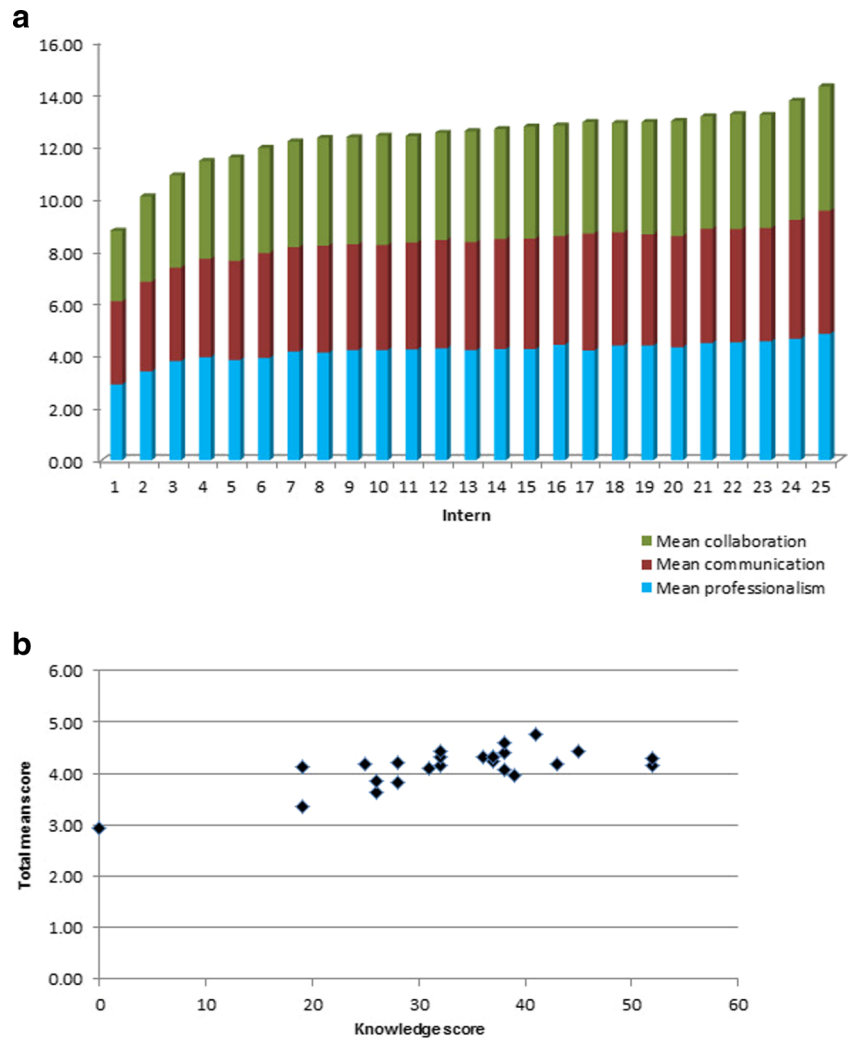
A comparison of average score given to each intern by different groups of raters is shown in Figs. 1b and 2.

Discussion

In a previous study we conducted in the BDF Hospital, we determined that our BDF/PCC instrument and the MSF process both were feasible, reliable, valid, and applicable to our setting. As such, we used it on this year’s cohort of interns for further assessment of their successful internship experiences and to provide more evidence to support the validity of the BDF/PCC instrument.

This study confirmed that the BDF/PCC instrument is feasible, reliable, and valid for the evaluation of professionalism, communication skills, and collaboration among clerkship physicians. The high response rates, number of minutes required to complete the questionnaire, and the small number of raters needed to produce a reliable assessment indicate the feasibility of the

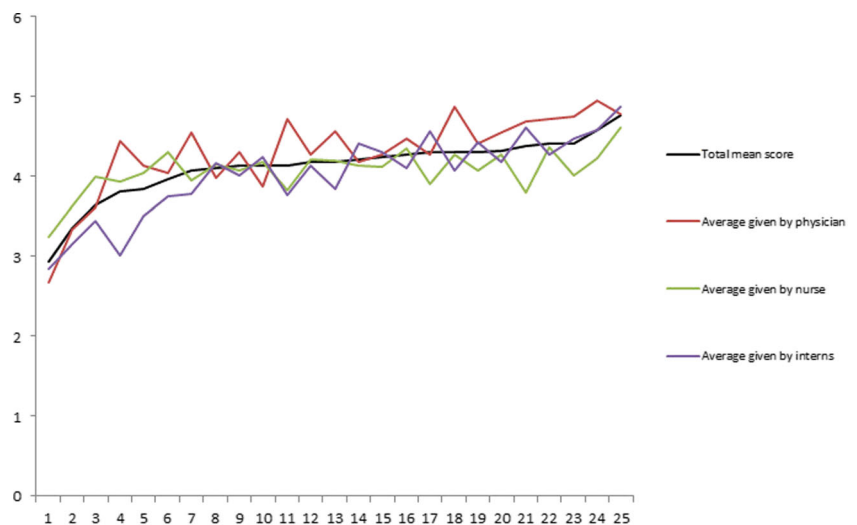
Fig. 1 a A stacked bar graph illustrating three stacks per bar to represent each of the mean score in professionalism, communication and collaboration per student and arrange student entries in ascending order of total mean score. **b** Line graph of total mean score and mean score of different raters



BDF/PCC instrument. The exploratory factor analysis in the previous study and the confirmatory factor analysis from this study resulted in three composite scales: professionalism,

communication skills, and collaboration. The factor analysis showed that the questionnaire could be grouped into three factors representing 77.3 % of the total variance.

Fig. 2 Total mean score given by physicians, nurses, and interns



Providing such feedback to physicians in their internship year is an essential part of the learning process [24]. Multisource feedback, also termed 360° evaluation, has become a popular method for assessing trainees in different fields [25]. However, because it has been documented that interns are oftentimes not observed sufficiently when conducting clinical activities in the field [24], it is critical that selected raters are those who have had the opportunity to truly observe the intern's performance [27]. Otherwise, not only will selecting the wrong raters lead to inaccurate ratings but also the interns themselves may not value or act upon the feedback provided by such raters, [28] thus rendering the MSF process inadequate.

Our study included a self-assessment tool as it has been established by past research that MSF questionnaires that include a self-assessment section have proven to be useful in further improving the learning process [29]. Self-assessment evaluations indicate the extent to which the individual is able to accurately self-assess—those who have a tendency to rate themselves higher than their supervisors would benefit from a constructive discussion to understand how to better monitor their own learning process. This is a lifelong skill that is essential to any field of expertise [30]. This discussion is an important one as they may ease the acceptability of potential criticism that may come as a result of the MSF study. There are several factors that affect an individual's acceptability of external feedback and whether they will use that feedback to improve themselves. These include distrust of the credibility of their raters, misunderstanding the feedback itself, and being uncertain of how to use the feedback to improve themselves, among others [30]. As such, a conversation between a facilitator and the intern after the results of the MSF have been distributed would be an important subsequent step following administration of the questionnaire in order to reap the most benefits of this tool.

The belief of the individual being assessed that the multisource feedback process is a credible and accurate means of self-improvement is an essential contributing factor to the likelihood of their making changes for improvement [31]. A study conducted on 113 family physicians found that 61 % made or planned to make changes in the way they practice as a direct result of the feedback they received via an MSF tool [32]. This study found that generally only those who felt that the process was credible and accurate are the ones who actively sought self-improvement based on the results of their feedback [33].

The major limitation of this study is our sample size, which was limited to 25 interns. Future studies may be useful to further examine the trends observed here with larger numbers and perhaps with other hospitals. Our study showed redundancy in some items, for the future use we will avoid these items for better evaluation. This study was an initial phase; for evaluating their skills improvement and interns' long lasting impression of the BDF/PCC instrument, we need to conduct another survey in the future in the same cohort of interns.

Conclusion

The BDF/PCC instrument to assess professionalism, communication, and collaboration skills seems to yield feasible, reliable, and valid measurements in assessing physicians in their clerkship year. Testing the instrument on two occasions with a 1-year interval can support the construct validity. It will be interesting to extend the implementation of the BDF/PCC instrument in other departments and specialties in the future.

Compliance with Ethical Standards

Ethical Approval The research was approved by the research ethics committee in the BDF hospital. Written consent was obtained from the interns, and verbal consent was obtained from raters. The study was conducted between March 2014 and June 2015.

Competing Interests The authors declare that they have no competing interests.

Financial Competing Interests All authors declare that there are no financial competing interests.

References

1. Archer JC, Norcini J, Davies HA. Use of SPRAT for peer review of paediatricians in training. *BMJ*. 2005;330:1251–3.
2. Al Ansari A, Donnon T, Al Khalifa K, Darwish A, Violato C. The construct and criterion validity of the multisource feedback process to assess physician and performance: a meta-analysis. *Advn in Med Educ & Reser*. 2014;5:39–51.
3. Moonen-van Loon JM, Overeem K, Govaerts MJ, Verhoeven BH, van der Vleuten CP, Driessen EW. The reliability of multisource feedback in competency-based assessment programs: the effects of multiple occasions and assessor groups. *Acad Med*. 2015;8:1093–9.
4. Wood L, Hassell A, Whitehouse A, Bullock A, Wall D. A literature review of multi-source feedback systems within and without health services, leading to 10 tips for their successful design. *Med Teach*. 2006;28:e185–91.
5. Druskatt V, Wolff S. Effects and timing of developmental peer appraisals in self-managing work groups. *J Appl Psychol*. 1999;1: 58–74.
6. Edwards M, Ewen A. 360 feedback: the powerful new model for employee assessment and performance improvement. New York: AMACOM; 1996.
7. Waldman D, Bowen D. The acceptability of 360° appraisals: a customer-supplier relationship perspective. *Hum Resour Manag*. 1998;2:117–29.
8. Van der Heijden BI, Nojhof AH. The value of subjectivity: problems and prospects for 360 degree appraisal systems. *Int J Hum Resour Manag*. 2004;3:493–511.
9. Brinkman WB, Geraghty SR, Lanphear BP, Khoury JC, del Rey Gonzalez JAG, DeWitt TG, Britto M. Effect of multisource feedback on resident communication skills and professionalism—a randomized controlled trial. *Arch Pediatr Adolesc Med*. 2007;1:44–9.
10. Garra G, Wackett A, Thode H. Feasibility and reliability of a multisource feedback tool for emergency medicine residents. *J Grad Med Educ*. 2011;3:356–60.

11. Donnon T, Al Ansari A, Al Alawi S, Violato C. The reliability, validity, and feasibility of multisource feedback physician assessment: a systematic review. *Acad Med.* 2014;3:1–6.
12. Al Ansari A, Al Khalifa K, Al Azzawi M, Al Amer R, Al Sharqi D, Al-Mansoor A, et al. Cross-cultural challenges for assessing medical professionalism among clerkship physicians in a middle eastern country (Bahrain): feasibility and psychometric properties of multisource feedback. *Advanc in Medi Educ and Pract.* 2015;6:509–15.
13. Lockyer JM, Violato C, Fidler H. The assessment of emergency physicians by a regulatory authority. *Acad Med.* 2006;12:1296–303.
14. Violato C, Lockyer JM, Fidler H. Assessment of psychiatrists in practice through multisource feedback. *Can J Psychiatr.* 2008;8: 525–33.
15. Van Thiel J, Van Dalen J, Ram P. MAAS Global Manual 2000. Maastricht: University Press; 2003.
16. Burt J, Elmore N, Campbell J, Roland M, Benson J, et al. Assessing communication quality of consultations in primary care: initial reliability of the global consultation rating scale, based on the Calgary-Cambridge guide to the medical interview. *BMJ Open.* 2014;4:e004339.
17. Archer J, Norcini J, Davies H. Use of SPRAT for peer review of pediatricians in training. *BMJ.* 2005;330:1251–3.
18. Orchard CA, King GA, Khalili H, Bezzina MB. Assessment of inter-professional team collaboration scale (AITCS): development and testing of the instrument. *J Contin Educ Heal Prof.* 2012;1:58–67.
19. Violato C, Saberton S. Assessing medical radiation technologists in practice: a multi-source feedback system for quality assurance. *Can J Med Radiat Technol.* 2006;2:10–7.
20. Lockyer J, Violato C, Fidler H, Alakija P. The assessment of pathologists/laboratory medicine physicians through a multisource feedback tool. *Arch Pathol Lab Med.* 2009;8:1301–8.
21. Renee L, Kiki L, Maas JH. Systematic evaluation of the teaching qualities of obstetrics and gynecology faculty: reliability and validity of the SETQ tool. *PLoS One.* 2011. 6(5).
22. Streiner DL, Norman GR. *Health measurement scales: a practical guide to their development and use.* Oxford: Oxford University Press; 2008.
23. Brennan RL. *Generalizability theory*, vol. 79. New York: Springer-Verlag; 2001. p. 441.
24. Shute VJ. Focus on formative feedback. *Rev Educ Res.* 2008;78:153–89.
25. Hawkins RE, Katsufraakis PJ, Holtman MC, Clauser BE. Assessment of medical professionalism: who, what, when, where, how, and ... why? *Med Teach.* 2009;31:348–61.
26. Pulito AR, Donnelly MB, Plymale M, Mentzer Jr RM. What do faculty observe of medical students' clinical performance? *Teach Learn Med.* 2006;18:99–104.
27. Mazor KM, Holtman MC, Shchukin Y, Mee J, Katsufraakis PJ. The relationship between direct observation, knowledge and feedback: results of a national survey. *Acad Med.* 2011;86:S63–8.
28. Dolmans DH, Wolfhagen IH, Heineman E, Scherpbier AJ. Factors adversely affecting student learning in the clinical learning environment: a student perspective. *Educ Health (Abingdon).* 2008;21:32.
29. Sargeant J. Reflecting upon multisource feedback as 'assessment for learning'. *Perspect Med Educ.* 2015;4:55–6.
30. Sargeant J, Mann K, van der Vleuten C, Metsemakers J. Reflection: a link between receiving and using assessment feedback. *Adv Health Sci Educ Theory Pract.* 2009;3:399–410.
31. Miller A, Archer J. Impact of workplace based assessment on doctors' education and performance: a systematic review. *BMJ.* 2010;341:c5064.
32. Sargeant JM, Mann KV, Ferrier SN, Langille DB, Muirhead PD, Hayes VM, et al. Responses of rural family physicians and their colleague and coworker raters to a multi-source feedback process: a pilot study. *Acad Med.* 2003;78:S42–4.
33. Sargeant J, Mann K, Ferrier S. Exploring family physicians' reactions to multisource feedback: perceptions of credibility and usefulness. *Med Educ.* 2005;39:497–504.