

Multisource feedback for educational supervisors: Validity evidence from the first year of implementation

Tim Swanwick MA (Cantab) MBBS FRCGP MA FAcadMED

Dean of Professional Development in the London Deanery, London, UK and Visiting Professor of Medical Education at the University of Bedfordshire, UK

Daniel Smith MSc

Research and Analytics Manager, London Deanery, London, UK

Catherine O’Keeffe EdD, MPhil, MEd, BNurs, RGN, RHV, DN cert

Professional Development Education Lead – Faculty Development, London Deanery, and Senior Lecturer in Clinical Education at the Institute of Education, University of London

Nerys Cater BA MA

Team Leader, Faculty and Leadership Development, London Deanery, London, UK

Julian Archer MBChB MRCPCH MEd PhD FAcadMED

National Institute for Health Research Career Development Fellow and Clinical Senior Lecturer in Clinical Education at the Peninsula College of Medicine and Dentistry, Plymouth, UK

Introduction

Previously we reported on the development of an unique online multisource feedback instrument for educational supervisors in postgraduate medical education (Archer et al 2012). Educational supervisors play a pivotal role in the training of doctors providing feedback and support for the learner in relation to their developmental trajectory while facilitating access to the resources of the training organisation and orchestrating appropriate learning experiences (Kilminster S, Cottrell D et al. 2007). The emerging ‘professionalisation’ of postgraduate medical education (Swanwick T 2008) coupled with increasingly explicit accountability of supervisors to employers, regulators and trainees, requires supervisors to evidence what they do, as well as demonstrating that they actively seek feedback on their own performance (London Deanery 2009). To enable this process, the London Deanery, an organisation responsible for around 20% of the UK’s postgraduate medical training, commissioned the development of a validated online multisource feedback (MSF) instrument in 2009. Following initial development and piloting, the instrument was introduced into the ‘live’ environment in July 2010.

In total 3,480 educational supervisors were identified and added into the MSF database during its first operational year; 97% of these cases were pre-loaded from a census of supervisors or from pre-existing deanery data. We then emailed all supervisors asking them to activate their accounts. This approach aimed to encourage take-up through minimising the data entry required by users to initiate the process. All supervisors received one reminder to take part during the year. Reports generated from the 18 item questionnaire provided feedback from trainees to supervisors on two aspects of their role; the degree to which they offered *challenge and support* and *personal attributes*, with a single item asking for an overall evaluation. Trainers were also invited to self-rate against the items.

Work in the developmental phase concentrated on providing evidence against a validity framework provided by Downing (Downing SM 2003). Data from the first year of implementation has provided further evidence for the validity of the instrument and has led to further refining and a programme of ongoing improvement. In addition to providing further support for the wider use of the multisource feedback instrument itself, this study raises some interesting general questions about the structure of multisource feedback, and how best to support supervisors in their role.

Methodology

Downing recommends that validity evidence for assessment should be collected under five headings; content, response process, internal structure, relationship to other variables and consequences. The mixed methodology adopted in this study enabled collection of data under all five headings. The methods used were as follows:

Quantitative descriptive statistics

Usage data for the first full year of implementation was collected together with rater response rates, numbers of items answered and frequency of ratings given.

Assessment of internal consistency (reliability)

Cronbach's alpha was used to assess the internal consistency of the scales previously identified, and under which the questionnaire was presented, namely; 'personal attributes' and 'challenge and support'. Alpha was also calculated for the instrument as a whole.

Assessment of construct validity

Construct validity evidence was provided through factor analysis using Principal Component Analysis with Varimax rotation. Kaiser normalization loadings lower than 0.32 were suppressed. N = 3,596

Relationships between scale scores and key demographic variables were explored. As the data were not normally distributed, a non-parametric approach was required for statistical testing. Each scale score was recoded so that 1 represented those in the top 25%. Logistic regression was then used to test whether a trainee's level of training (their grade) and the types of supervision they received were associated with their scores being in the top 25% of scores.

Exploration of content validity through qualitative analysis of free text responses

A random sample of 15% of the first 405 educational supervisors to have completed the multisource feedback were selected resulting in 62 cases to be analysed. All the free text trainee comments pertaining to these cases together with supervisor self-assessments were reviewed and a coding frame developed based on the MSF instrument's scale categories other comments. Content was coded in each relevant category and recurring themes identified.

Exploration of consequential validity through a qualitative study of perceptions of educational supervisors

We undertook semi-structured telephone interviews to elicit the perceptions of educational supervisors who had received feedback. A non-probability sampling method (Cohen L, Manion L et al. 2000), was adopted and included only those who had completed the instrument and received the feedback report within the 6 months leading up to data collection. 45% of the sample population (n = 115) volunteered to participate in the study. As responses were received from across the specialties it was decided to purposefully sample the 25 participants from across 5 individual specialty groups identified; general practice (n=1), surgical specialties (n = 3), anaesthesia (n = 4); medical specialties (n =4), paediatrics (n =6) and psychiatric specialties (n = 6).

Questions for the interview guide were developed using key themes identified from a literature review using standard journal databases with key search terms 'multisource', 'feedback' and 'supervision'. Themes included; motivations for undertaking the instrument, perceptions of accuracy and usefulness of the feedback, whether areas for improvement had been identified and if so, whether these had been implemented in practice along with the usefulness of discussing the feedback report generated by the instrument. Draft interview guides were developed then checked for content validity along for any potential sensitive issues. All interviews were approximately thirty minutes in duration, recorded using a digital telephone recorder and were transcribed verbatim.

Results

Quantitative descriptive statistics

Of the 3,480 educational supervisors in the database, 1,207 (35%) had started the process by the end of the first full year of implementation (July 2010-June 2011). Of these, 665 had completed the process and received a report with 19 supervisors having completed the process twice. For those that completed the process, the mean response rate from nominated trainees was 82%. Tables 1 and 2 summarise the responses from trainees drawn from completed reports. It can be seen that the data are highly skewed i.e. for all items, 98% of the ratings are at satisfactory or above. Table 1 maps the items to the scales developed from the pilot data.

Table 1

Trainee ratings

Item	Scale	N	N unable to comment	% unable to comment	Mean score (out of 6)	Standard deviation
Ability to inspire you	Challenge and support	3577	19	0.5%	5.48	0.73
Ability to challenge you	Challenge and support	3567	29	0.8%	5.47	0.66
Willingness to act to resolve problems in a timely manner	Challenge and support	3542	54	1.5%	5.60	0.65
Ability to give constructive feedback	Challenge and support	3574	22	0.6%	5.54	0.68
Encouragement towards you achieving excellence	Challenge and support	3573	23	0.6%	5.56	0.68
Ability to take your supervision beyond a tick box exercise	Challenge and support	3543	53	1.5%	5.56	0.68
Ability to offer practical tailored advice for your longer term career planning	Challenge and support	3470	126	3.5%	5.45	0.74
Ability to remain up-to-date about your training scheme	Personal attributes	3547	49	1.4%	5.41	0.70
Genuine interest in your portfolio	Personal attributes	3481	115	3.2%	5.36	0.74
Approachability	Personal attributes	3591	5	0.1%	5.73	0.57
Enthusiasm	Personal attributes	3588	8	0.2%	5.65	0.60
Communication skills	Personal attributes	3589	7	0.2%	5.63	0.63
Honesty & Integrity	Personal attributes	3568	28	0.8%	5.76	0.51
Ability to assure privacy and where appropriate confidentiality	Personal attributes	3490	106	2.9%	5.67	0.58
Ability to make time for you	Personal attributes	3575	21	0.6%	5.59	0.66
Interest in you as an individual	Personal attributes	3575	21	0.6%	5.55	0.68
Ability to be your advocate	Personal attributes	3375	221	6.1%	5.51	0.71
Overall, record a global judgement compared with other colleagues	Not included in scale or Overall Mean	3587	9	0.3%	5.69	0.59
<i>Total responses</i>	<i>3596</i>					

Table 2

Response frequencies

Item	Very Poor (%)	Poor (%)	Needs development (%)	Satisfactory (%)	Good (%)	Very Good (%)	N excluding unable to comment
Ability to inspire you	0.1	0.1	1.3	8.9	29.4	60.2	3,577
Ability to challenge you	0.0	0.1	0.6	6.8	37.5	55.0	3,567
Willingness to act to resolve problems in a timely manner	0.1	0.1	0.5	5.7	26.2	67.4	3,542
Ability to give constructive feedback	0.1	0.1	0.9	6.4	29.4	63.1	3,574
Encouragement towards you achieving excellence	0.1	0.1	0.9	6.5	27.6	64.8	3,573
Ability to take your supervision beyond a tick box exercise	0.1	0.2	0.7	6.9	27.1	65.0	3,543
Ability to offer practical tailored advice for your longer term career planning	0.1	0.1	1.1	9.9	31.0	57.8	3,470
Ability to remain up-to-date about your training scheme	0.0	0.1	0.6	10.1	36.8	52.4	3,547
Genuine interest in your portfolio	0.1	0.2	0.9	11.6	37.0	50.2	3,596
Approachability	0.1	0.1	0.4	3.9	17.0	78.6	3,591
Enthusiasm	0.0		0.4	5.3	23.3	71.1	3,588
Communication skills	0.0		0.7	5.6	24.0	69.7	3,589
Honesty & Integrity	0.0	0.0	0.1	3.1	17.0	79.7	3,568
Ability to assure privacy and where appropriate confidentiality	0.1	0.1	0.2	4.4	22.7	72.6	3,490
Ability to make time for you	0.1	0.2	0.6	6.0	26.3	66.9	3,575
Interest in you as an individual	0.1	0.2	0.6	6.9	28.0	64.2	3,575
Ability to be your advocate	0.1	0.2	0.8	8.3	29.2	61.5	3,375
Overall, record a global judgement compared with other colleagues	0.1	0.0	0.6	4.3	20.7	74.2	3, 587

Assessment of internal consistency (reliability)

Cronbach's alpha for the instrument as a whole (18 items) $\alpha = 0.95$ (N = 3,123)

Cronbach's alpha for the scale 'challenge and support' (7 items) $\alpha = 0.90$ (N = 3,390)

Mean scale score calculated if 5 were answered. 'Unable to comment' excluded to reduce missing data.

Cronbach's alpha for the scale 'personal attributes' (10 items) $\alpha = 0.91$ (N = 3,217)

Mean scale score calculated if 9 were answered.

This level of internal consistency is considered good and in fact there may be some redundancy inferring that in future some items could be removed without affecting reproducibility (Nunnally J 1978).

Assessment of construct validity

Factor analysis of the items did not reproduce the two-factor solution found with the pilot data [REF]. Only one factor with an Eigen value greater than 1 emerged accounting for 55% of the variance (N = 3,124). This is unusual as MSFs typically factor analyse into a two-factor solution an 'interpersonal or relationship' factor, and a 'knowledge or skill' factor' (Wood L, Hasseell A et al. 2006). As this has implications for the construct validity of the instrument, we explored this unexpected result further and in order to eliminate possible response bias - where, for instance, a trainee might be unwilling to offer negative feedback - experimented with scoring 'unable to comment' as 0 as if the trainee had experienced educational supervision in full they should have been able to comment. The result was a more interesting three factor-structure with the emergence of a third factor with items that appear specifically 'educational', such as "Genuine interest in your portfolio". See Table 3.

Elsewhere in this analysis we have continued to use the scales developed from the pilot work as this was how the results were reported by the instrument to the participants.

The effect of trainee grade and type of supervision provided on ratings was examined and as Table 4 shows there were small differences. For both scale scores this was the case (Table 5) - trainees in higher grades gave higher scores and trainees receiving both clinical and educational supervision gave higher scores. This pattern is consistent with the reported literature (Kilminster S, Cottrell D et al. 2007) and suggests that the scales are capturing relevant variance, strengthening the nomological net. (Cronbach L and Meehl P 1955) Trainees in high grades also give higher rating on the UK's National Survey of Trainee Doctors. (Smith D, Riley S et al. 2007)

Table 3

Factor structure with unable to comment scored zero

	Challenge	Interpersonal	Educational
Percentage variance explained	45%	7%	6%
Ability to challenge you	0.76		
Ability to inspire you	0.74		
Encouragement towards you achieving excellence	0.70		
Ability to give constructive feedback	0.63	0.34	
Ability to take your supervision beyond a tick box exercise	0.55		0.49
Willingness to act to resolve problems in a timely manner	0.37	0.35	
Honesty & Integrity		0.75	
Communication skills	0.42	0.71	
Approachability	0.38	0.70	
Enthusiasm	0.49	0.61	
Ability to assure privacy and where appropriate confidentiality		0.58	0.39
Interest in you as an individual	0.38	0.48	0.47
Genuine interest in your portfolio			0.77
Ability to remain up-to-date about your training scheme			0.66
Ability to be your advocate			0.61
Ability to offer practical tailored advice for your longer term career planning	0.41		0.61
Ability to make time for you		0.50	0.52
<i>Extraction Method: Principal Component Analysis.</i>			
<i>Rotation Method: Varimax with Kaiser Normalization. Loadings lower than 0.32 are suppressed.</i>			
<i>N = 3,596</i>			

Table 4

Mean scale scores by type of supervision and trainee grade

Grade of trainee rater	Type of supervision received		Challenge and Support	Personal attributes
Foundation	Educational Supervision only	Mean	5.36	5.53
		N	76	74
	Educational and Clinical Supervision	Mean	5.51	5.58
		N	449	443
	Total	Mean	5.49	5.58
		N	525	517
Core	Educational Supervision only	Mean	5.37	5.51
		N	148	143
	Educational and Clinical Supervision	Mean	5.52	5.59
		N	989	966
	Total	Mean	5.50	5.58
		N	1137	1109
Higher	Educational Supervision only	Mean	5.39	5.40
		N	92	91
	Educational and Clinical Supervision	Mean	5.60	5.63
		N	1047	1029
	Total	Mean	5.58	5.62
		N	1139	1120
Total	Educational Supervision only	Mean	5.37	5.48
		N	316	308
	Educational and Clinical Supervision	Mean	5.55	5.61
		N	2485	2438
	Total	Mean	5.53	5.59
		N	2801	2746

Table 5

Type of trainee and type of supervision and their relationship to the scale scores

		B	S.E.	Wald	Sig.	Odds ratio	95% C.I. for Odds Ratio		Overall Model
							Lower	Upper	
Challenge and Support	Type of supervision received		.142	7.471	.006	1.472	1.116	1.943	$X^2 = 37.61$, P < 0.001
	Grade of trainee rater	.297	.058	26.590	.000	1.346	1.202	1.506	
	Constant	-1.862	.182	104.588	.000	.155			
Personal attributes	Type of supervision received	.306	.141	4.721	.030	1.358	1.030	1.789	$X^2 = 12.12$, P = 0.002
	Grade of trainee rater	.143	.058	6.177	.013	1.154	1.031	1.291	
	Constant	-1.519	.180	71.598	.000	.219			

Trainer variables and rating scores

There was no relationship between supervisors' gender, and supervisors' self-rating or the mean trainee rating scores on the two pre-existing scales. There were very small correlations between supervisors' year of qualification and mean trainee rating scores on the two scales. (Challenge and support Spearman's $r = 0.096$, $P = 0.013$, Personal attributes $r = 0.07$, $P = 0.0225$); showing that qualifying more recently was marginally associated with higher scores.

Supervisors' self-ratings were lower than the mean of the ratings from trainees (see Table 6).

Table 6

Trainer self-ratings compared to the mean of ratings from trainees

Scale	Minimum	Maximum	Mean	Lower 95% CI	Upper 95% CI	Std. Deviation
Challenge and Support - self	2.86	6.00	4.75	4.71	4.80	0.57
Challenge and Support - trainee	4.30	6.00	5.52	5.50	5.54	0.30
Overall - self	3.00	6.00	4.87	4.83	4.91	0.52
Overall - trainee	4.14	6.00	5.56	5.54	5.58	0.28
Personal Attributes - self	3.10	6.00	4.95	4.91	4.99	0.54
Personal Attributes - trainee	4.23	6.00	5.59	5.57	5.61	0.27

There were relationships between qualifying in the UK and the scale scores: UK-qualified supervisors had higher mean ratings from trainees; but lower self-ratings than supervisors qualifying outside of the UK. The scores and P values are reported in Table 7.

Table 7

Trainee and self ratings by supervisor's place of qualification

Supervisor's country of qualification		Challenge and Support - trainee mean	Personal attributes - trainee mean	Overall - trainee mean	Challenge and Support - self	Personal attributes - self	Overall - self
Not UK qualified	Mean	5.45	5.53	5.49	4.89	5.07	4.99
	N	196	196	196	196	196	196
	Std. Deviation	0.33	0.29	0.29	0.57	0.56	0.54
UK qualified	Mean	5.55	5.61	5.59	4.70	4.90	4.81
	N	469	469	469	469	468	468
	Std. Deviation	0.29	0.26	0.26	0.56	0.53	0.50
Total	Mean	5.52	5.59	5.56	4.75	4.95	4.87
	N	665	665	665	665	664	664
	Std. Deviation	0.30	0.27	0.28	0.57	0.54	0.52
Mann-Whitney Test P value		<0.001	<0.001	<0.001	<0.001	<0.001	<0.001

Supervisors who qualified outside the UK are more likely to give themselves higher ratings than their trainees (Table 8): 21% of non UK-qualified supervisors had higher self-overall score compared to the mean overall score from their trainees; compared to 8% from UK-qualified supervisors. The overall-score mean is the mean of the 17 items and does not include the global judgement item.

Table 8

Supervisor's place of qualification and self rating compared to trainees' overall mean rating (overall)

Supervisor's country of qualification	Overall self below trainee ratings	Overall self above trainee ratings	N
Not UK-qualified	79%	21%	196
UK qualified	92%	8%	469
Total	88%	12%	665
$\chi^2 = 22.40, P < 0.001$			

Qualitative analysis of free text responses

Overall the comments made by trainees about the educational supervisors were very positive. 269 out of 319 trainees commented on positive challenge and support of which approachability (n = 62), providing careers advice (n = 57), and being inspiring (n= 51) were the most frequently occurring:

very approachable, listens to my needs, takes time to understand them and helps out in every way she can. I feel her knowledge is something I hope to attain/achieve, and feel comfortable approaching Dr X regarding any matters related to this (101)

In addition clinical knowledge and expertise were commented on (n=32):

This person has the fascinating ability of being inspiring and challenging at all times. His clinical practice is a permanent example to be followed and his neurosurgical knowledge extensive and supported by a wide clinical experience. (301)

With regard to personal attributes, the themes that emerged were similar to those noted under the positive support and challenge category, with being approachable commented on most frequently, followed by enthusiasm and empathy.

Only 27 trainees made specific comments on negative support and challenge. Of these, 8 comments in total related to giving feedback include ability to be forthcoming with feedback and ability to give constructive criticism. Examples include:

I have not had the chance to have a formal meeting with Dr Y which would be helpful in forming a plan for my period of training with her and to allow her to give me some practical career advice. I would also appreciate the use of the staged approach to teaching/learning to help me progress in the more complex procedures that I am starting to undertake. (15)

Two comments suggested possible cause for concern:

Positive advice and support that isn't followed through when needed. At time of need, feel unsupported and backstabbed. (310)

Positive support sometimes lacking at times of difficulty. Did not feel safe. (139)

13 trainees made specific comments in relation to negative personal attributes of trainers of which 7 referred to communication and feedback issues:

Not necessarily interested in being your advocate-for example with a patient complaint I had, she was more focused on things I did wrong vs what I did right. Needs to be more balanced with feedback. (203)

When considering the self reflections of the trainers, 36 out of 62 trainers made comments in relation to providing personal support and challenge, 9 made comments related to taking an interest in the trainee:

I have taken time to listen to trainees and try as much as possible to be supportive and developmental in my approach. (26)

Nine comments related to providing careers advice:

I have helped number of trainees with career planning and placement (31)

Other less frequently occurring comments were made on the importance of good communication as the basis of effective feedback; understanding of and making time for the assessment process and the importance of encouraging reflection.

27 out of the 62 trainers made comments related to positive personal attributes.

14 comments related to being approachable:

I have had positive feedback from trainees about my interest, integrity and approachability (Trainer 7)

13 commented on their concern for, and interest in the trainees:

I will do everything I can to ensure I have time for my trainees. When unexpected problems have needed addressing I have ensured I have made time for a trainee to address concerns. (Trainer 56)

10 comments related to making time for the educational and assessment process:

Trainees are able to come into my office unannounced to speak to me at any time and they frequently do. I always ask trainees if they have time to complete their workload. I always review their assessments. (11)

33 out of 62 trainers commented on negative support and challenge. 10 comments related to difficulties giving feedback effectively:

I am not very patient and sometimes struggle with the time it takes to go through difficult situations. I sometimes see the obvious negatives without stressing the positives first. (26)

10 trainers made comments relating to negative personal attributes'. 7 comments related to workload and lack of time for educational supervision:

With an ever expanding workload it has been difficult to find time for trainee appraisal. (36)

8 comments related to lack of time:

Lack of time to be able to take educational supervision beyond tick boxing is frustrating (17)

Qualitative study of perceptions of educational supervisors

The individual interviews with educational supervisors who had completed the instrument highlighted their motivation for engaging with the MSF process, their views on the validity of the instrument and how they tried to improve their practice in response to feedback from trainees via the MSF.

The main reason given for deciding to use the instrument was to gain evidence of competence for educational appraisal portfolios and in some cases both educational and clinical appraisals.

"I'm trying to keep my educational portfolio up to date so I thought that it would be very helpful for that. It's also very useful as part of my hospital appraisal process as well" (INT 2)

The supervisors interviewed however expressed a genuine interest in wanting to discover how they were performing in the role and identifying whether there were any areas needing improvement.

"...the reason that I wanted feedback is that I want to be doing more of what trainees find useful and less of what they find unhelpful, so making more efficient use of my time... if you don't get feedback you are never quite sure of where you are pitching things." (INT5)

All the educational supervisors interviewed noted that the positive feedback received from trainees was both affirming and motivated them to further engage with the role.

I think it was very positive actually much more than which I expected. So in that way I think I have learned a lot and that has given me some encouragement as well to carry on with the supervision (INT 18)

Doubts were expressed however regarding the validity of the instrument if trainees are concerned that anonymity cannot truly be assured when educational supervisors work with small numbers of trainees. The supervisor commented that such concerns may prevent trainees from being critical when completing the instrument

Interviewee: I think ... trainees find it very difficult to ... give feedback to their consultants in this format.

Interviewer: Even though it was anonymous?

Interviewee: Yes. Because it's a limited number. (INT 2)

Educational Supervisors were able to identify areas for improvement as a result of the comments that they received including ensuring protected, uninterrupted time for educational supervision sessions, and using a more formal, structured approach to clinical teaching.

Discussion

This study has provided a detailed analysis of a substantial quantity of real data about a multisource feedback instrument in use in the live environment. The fact that it has been conducted 'in the field' adds to the validity of the study, although some operational changes in response to user feedback were made during the first year of use which may have impacted on the data (see footnote¹). Despite this, the 18 item instrument appears reliable, items perform well although are highly skewed, the content appears robust and there is evidence of (convergent) construct validity. Factor analysis has raised some interesting questions about the instrument itself and multisource feedback in general which warrant further exploration, perhaps by asking why respondents were unable to comment on the instrument itself.

The response process seems effective although issues about anonymity continue to cause concern. A paucity of constructive free-text feedback from trainees raises questions about how to enhance this and in order to provide specific and actionable feedback. Subsequent development of the instrument requires

¹ Two changes were made during the first year in response to user feedback: (i) the threshold to receive a report was lowered with a report available to supervisors on request with just 3 trainee respondents accompanied by a warning about reliability. (ii) additional validation checks were put in place to ensure that nominated raters identified the type of supervisory relationship they had with their supervisor. Those receiving purely clinical supervision were excluded from the process.

that all supervisees select one area for their supervisor's development, regardless of how highly they have rated the supervisor generally. The list of areas is populated with any item from the 17 on the instrument rated less than 'very good'. A mirror item has been included on the self-assessment.

Supervisors who engaged with the instrument appear to have a genuine interest in the role and in trying to help trainees learn. Supervisors found the instrument useful in a number of ways. They have understanding of role and self-awareness, but lack confidence – this instrument appears to be helpful in affirming what they do, which is important in the newly reified territory of supervision. A couple of comments about safe clinical practice raise the issue that many educational supervisors are also functioning as clinical supervisors, or at least need to maintain some oversight on clinical practice. This has led us to revisit some of the questions initially discarded after the pilot study. Subject to further validation of content validity, the instrument will be extended to all those with a formally recognised training role i.e. both named educational (programme) and named clinical (placement) supervisors.

Conclusion

Given that the multisource feedback instrument for supervisors is the only formal way for trainees to feed back on their supervisor as an individual, it is essential that they are not inhibited to provide accurate feedback. After all, educational supervisors in this study completed the instrument in order to seek an assessment on how they are performing in the role and whether there are any areas needing improvement. If trainees do not provide constructive feedback for improvement then this will impact on the effectiveness of the instrument and its contribution towards professional development. If trainee feedback is an integral part of the quality assurance process in clinical education, as is the focus in today's higher education sector (Department for Business 2011) it will be essential to provide guidance and support for trainees on providing accurate and constructive feedback to supervisors in order for them to contribute towards maintaining the standards of their training. Indeed this could be argued to be a professional duty. There is a significant amount of literature on effective feedback from trainer to trainee. One of the issues that this study has highlighted is that there needs to be more guidance for those who receive the training and the feedback that they can provide to the trainers.

References

Archer J, Swanwick T and Smith D (2012) Developing a multisource feedback tool for postgraduate medical education supervisors.

Online. Available at: <http://www.faculty.londondeanery.ac.uk/supervisor-MSF> (last accessed 30.5.12)

Downing SM. Validity: on the meaningful interpretation of assessment data. *Medical Education*. 2003;**37**:830-837.

Cohen L, Manion L, K. M. *Research methods in education*. 5th ed. London RoutledgeFalmer; 2000.

Cronbach L, Meehl P. Construct validity in psychological tests. *Psychological Bulletin*. 1955;**52**:281-302.

Department for Business IaS. *Students at the Heart of the System*. London Department for Business, Innovation and Skills; 2011.

London Deanery. *Professional Development Framework for Supervisors*.

<http://www.faculty.londondeanery.ac.uk/professional-development-framework-for-supervisors> Accessed 18th November 2011; 2009.

Kilminster S, Cottrell D, Grant J, B. J. AMEE Guide No. 27: Effective educational and clinical supervision. *Medical Teacher*. 2007;**29**:2 - 19.

Nunnally J. *Psychometric Theory*. New York McGraw Hill; 1978.

Smith D, Riley S, Kazmierczak A, Aitken M, Paice E, Le Rolland P. *National Survey of Trainees Summary Report* London Postgraduate Medical Education and Training Board; 2007

Swanwick T. See one, do one, then what? Faculty development in postgraduate medical education *Postgraduate Medical Journal* 2008;**84**:339-343.

Wood L, Hasseell 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. *Medical Teacher*. 2006;**28**:185-191.

Acknowledgements

Thanks to Michael Brett from WASP software for providing and maintaining the platform through which the MSF is administered and to Ruhi Merali who assisted in the collation of the data.

Declarations of interest

Research and development funded by the London Deanery.

Ethical approval

As a 'service evaluation' the NHS National Research Ethics Service does not require specific ethical approval, however, ethical approval for the qualitative aspects of the study was obtained through the Institute of Education, London, UK as a supervised Masters' research project.

