A multisource feedback tool to assess ward round leadership skills of senior paediatric trainees: (2) Testing reliability and practicability

Helen M Goodyear, Indumathy Lakshminarayana, David Wall, Taruna Bindal

ABSTRACT

Background A five-domain multisource feedback (MSF) tool was previously developed in 2009–2010 by the authors to assess senior paediatric trainees’ ward round leadership skills.

Objectives To determine whether this MSF tool is practicable and reliable, whether individuals’ feedback varies over time and trainees’ views of the tool.

Methods The MSF tool was piloted (April–July 2011) and field tested (September 2011–February 2013) with senior paediatric trainees. A focus group held at the end of field testing obtained trainees’ views of the tool.

Results In field testing, 96/115 (84%) trainees returned 633 individual assessments from three different ward rounds over 18 months. The MSF tool had high reliability (Cronbach’s $\alpha$ 0.84, G coefficient 0.8 for three raters). In all five domains, data were shifted to the right with scores of 3 (good) and 4 (excellent). Consultants gave significantly lower scores ($p<0.001$), as did trainees for self-assessment ($p<0.001$). There was no significant change in MSF scores over 18 months but comments showed that trainees’ performance improved. Trainees valued these comments and the MSF tool but had concerns about time taken for feedback and confusion about tool use and the paediatric assessment strategy.

Conclusions A five-domain MSF tool was found to be reliable on pilot and field testing, practicable to use and liked by trainees. Comments on performance were more helpful than scores in giving trainees feedback.

INTRODUCTION

In a multidisciplinary team it is usually the doctor who leads the ward round. There is, however, large variability in the way ward rounds are conducted, which can have a profound effect on the well-being of patients and medical staff. A number of attributes are required to lead a ward round and these have been explored in detail in a companion paper.

Ward round supervision has been found to be lacking. When rated on a 5-point scale (where 1 = not at all and 5 = almost always), the mean score of senior trainees was only 2.9. Thus, ward round supervision is not happening on a sufficiently regular basis. Without supervision and feedback, errors and deficits are perpetuated. This is of concern as ward round studies have found failure to perform the basics of patient care with key deficits in physical examination, reviewing charts (case notes and vital observations), prescriptions and documentation. Particularly worrying is the finding that physical examination by upper level residents (senior trainees) in approximately half the patients was brief, superficial and accounted for less than 10% of ward round time. Vital signs, medications and notes were not reviewed.

In a companion paper we describe the development of a five-domain multisource feedback (MSF) tool to investigate ward round leadership skills. Although workplace-based assessments have become an integral part of specialty training in the UK and internationally, some still view them as a tick box exercise. It is therefore important to design any new tool so that it is a valuable addition to existing assessments rather than a burden to trainees.

This study had the following objectives:

1. To pilot test the use of the MSF tool described in the companion paper.
2. To field test the tool with senior paediatric trainees in the West Midlands to see if it is practicable and reliable tool.
3. To see if there was variation in individuals’ feedback from the MSF ward round leadership tool over time.
4. To seek trainees’ views of the MSF tool.

METHODS

Piloting

A pilot study was undertaken in April–July 2011 with 20 trainees working at a large university hospital in our region to look at reliability and practicability. We had suggested having at least three assessors to complete the forms and it was essential to determine whether this gave reliable results. Additionally, it was important to see if the MSF tool could be used with ease as part of normal clinical practice or if there were problems with its use.

Field testing

Participants

Between September 2011 and February 2013 the MSF was used by senior paediatric trainees in the West Midlands to assess ward rounds that they were leading. Sampling was purposive, with an invitation sent to all 115 senior trainees working in the West Midlands region—that is, specialist registrars (SpRs) and those in ST4–8 (training years 4–8; see box 1 for UK paediatric training structure); trainees on maternity leave and out of programme were excluded. The trainees were invited to complete the MSF tool as an optional workplace-based assessment and evidence of competency in ward round leadership. They were reminded every 6 months at regional study days that this MSF tool was available for completion.

was accepted with a p value of 0.10. A general linear mixed model was performed measuring essentially the same construct and some items could indicate a high level of item redundancy with all items being unnecessary.

The reliability of the MSF tool was calculated using Cronbach’s α. A value of 0.7 is generally regarded as acceptable for a research study and a value of 0.8 for a high stakes assessment,9 but probably not higher than 0.9 because, if α is very high, it might indicate a high level of item redundancy with all items measuring essentially the same construct and some items could be unnecessary.10 A general linear mixed model was performed to test whether differences in MSF tool scores given by assessors were statistically significant. The total score was the dependent variable, assessor category was the fixed factor and whether first, second or third MSF and senior/junior trainee status were random factors. Levene’s and Hartley’s F max variance tests were used to look at whether the error variance of the dependent variable is equal across groups.11

A generalisability (D) study of both the pilot and field testing data was undertaken using GENOVA V3.1 with the number of domains held constant (the five domains in the MSF tool) and the numbers of raters varied.9 This was to look at how many raters were needed to get consistent ratings (G coefficient ≥0.8).

The results were also analysed to look for differences between self-assessment and those of the assessors and to establish how the scores varied over time for individuals.

Qualitative data
The focus group data were analysed thematically using principles described by Coffey and Atkinson.12 The detailed methodology is described in the companion paper (see box 2).2

**RESULTS**

Piloting

Practicability
Pilot respondents included a range of medical and nursing staff who reported no difficulties in using the MSF tool. Entry of data into the database was straightforward.

Reliability
Twenty senior trainees took part in the pilot. The results from 16 trainees had sufficient data (at least three complete assessments for all domains excluding self-assessment) for statistical analysis using a generalisability D (decision) study to be carried out. The G coefficients were 0.73, 0.8 and 0.84 for two, three and four assessors, respectively, showing that reliable results were obtained with three assessors.

Field testing
Ninety-six of the 115 trainees (83%) took part in field testing and completed 633 individual assessments (536 by assessors and 97 self-assessments; tables 1 and 2). The numbers of
assessments were 303, 210 and 120 assessments in rounds 1, 2 and 3, respectively. Reliability for the MSF tool was high (Cronbach’s α 0.84), confirming the pilot results. Just over half the trainee assessments (54%) were for those in level 2 training (ST4–5).

MSF tool scores
Overall the scores were high for all five domains with scores of 3 (good) and 4 (excellent) predominating (table 3). Low scores of 1 (needs improvement) and 2 (borderline) were given for preparation and organisation (18 assessments), communication skills (7 assessments), teaching and enthusiasm (32 assessments) and team working (5 assessments).

Comparison of scores given by different assessor groups
There were highly significant differences in scores in all five domains and overall scores by general linear mixed model (p<0.001; tables 4 and 5). Levene’s test for equality of error variances showed F=4.19, DF1=23 and DF2=420 (p<0.001). However, with large sample sizes, small differences in group variances can produce a Levene’s test that is significant because the power of the test is improved. Comparing the highest and lowest variances (SDs squared) and dividing the smallest value into the largest value gives a value of 2.44. Using Hartley’s F Max variance test this looks to be non-significant, so the result of Levene’s test can be disregarded.11 Senior doctors (consultants and staff grade and associate specialist (SAS) doctors); trainee doctors (all trainees including foundation doctors and associate registrars); nurses; and senior nurses (senior nurses and advanced nurse practitioners).

Self-assessment versus that of the assessors
Self-assessment scores were significantly lower (mean 15.3) in all five domains than assessors’ scores (p<0.001; table 4). This was also evident from looking at comments illustrated by the following case.

Example case
Trainee self-assessment comment: “Some investigations not checked before seeing patient. Some significant delays. Need to be more assertive with team to improve efficiency of ward round. Ward round a little disjointed and junior members not consistently there.”

Assessor comments: “Excellent team player, always feel well supported. Dr X is always calm and polite. Very busy ward round; despite this, Dr X organised the round in a very time efficient manner. Speaks to parents at appropriate level. Encourages questions to help their understanding. Supportive of nursing staff. Responds to their concerns regarding patient condition or concern.”

Comparing scores of junior and senior trainees
Trainees’ assessments were grouped into two variables: junior middle grade trainees (ST3 and ST4) with 244 assessments and senior trainees (ST5–8 and SpRs) with 389 assessments. There were no significant differences by the Mann–Whitney test and general linear mixed model.

Comparing scores from MSF 1, 2 and 3
Scores were compared in consecutive rounds to see if performance improved. There were no significant differences in any of the five domains over an 18-month period. However, free text comment showed that improvement was occurring, as illustrated by the following case.
Theme 1: Observation of practice

There was strong feeling that the MSF tool facilitated trainees being observed leading ward rounds which had not previously been the case in all units.

**Table 5** General linear mixed model testing of differences in scoring between assessor categories

<table>
<thead>
<tr>
<th>Assessor categories*</th>
<th>p Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>MSF 1, 2, 3</td>
<td>0.01</td>
</tr>
<tr>
<td>Junior/senior trainee status</td>
<td>0.71</td>
</tr>
<tr>
<td>Assessor categories×MSF 1, 2, 3</td>
<td>0.39</td>
</tr>
<tr>
<td>Assessor categories×junior/senior trainees</td>
<td>0.39</td>
</tr>
<tr>
<td>MSF 1, 2, 3×senior/junior trainee status</td>
<td>0.51</td>
</tr>
<tr>
<td>Assessor categories×MSF 1, 2, 3×junior/senior trainee status</td>
<td>0.55</td>
</tr>
</tbody>
</table>

*Assessors were grouped into four categories: senior doctors (consultants and staff grade and associate specialist (SAS) doctors); trainee doctors (all trainees including foundation doctors and specialty registrars); nurses; and senior nurses (senior nurses and advanced nurse practitioners).

**MSF 1**

Assessor comments: “Pace needs to be quicker without compromising communication. Not much teaching. Mostly business round. Discussion briefly on ECG and management of Kawasaki’s disease.”

Self-assessment comments: “I did not teach as much as I usually do. Needs improvement. The ward round continued for too long in my opinion. Needs improvement.”

**MSF 2**

Assessor comments: “Improvement noted in organisation and pace. Busy business round but still taught where able. Better time management compared with previous MSF despite ward being very busy.”

Self-assessment comments: “I feel I improved with organisation and time effectiveness. I made an effort to teach during the assessment as I would normally.”

Focus group comments

Theme 2: Speed of feedback

This had been slow due to the MSF forms being collated by the postgraduate administrator centrally and could be a month or longer. Some forms had been lost in the post. It was felt that educational supervisors would be able to undertake this task much more speedily and enable feedback to be much quicker.

Theme 3: Scoring system

The scores were felt not to be helpful as they did not seem to change. Comments were highly appreciated and had modified practice.

Theme 4: Confusion about the number of assessments

Trainees were confused about whether this was an addition to the mandatory Royal College of Paediatrics and Child Health assessments or instead of some of them. They were also uncertain about how many to do in 6 months.

Theme 5: Not practicable in all specialties

In some subspecialties such as community paediatrics and in the emergency department ward rounds did not take place and these trainees were worried that they would fail their annual review of competence progression (ARCP) due to being unable to complete the MSF tool.

Theme 6: Concern that this was a summative assessment

Trainees were concerned that this was a pass/fail assessment for ARCP and had been worried about undertaking it. As stated in theme 2, they welcomed comments.

**DISCUSSION**

In this study a five-domain MSF tool on ward round leadership for senior trainees was practicable and reliable on pilot and field testing. It was of interest that there was no significant change in individuals’ MSF scores over an 18-month period, in contrast to assessors’ comments which showed improved trainee performance. It is also of interest that trainees valued the tool but found comments more helpful than the scoring system.

As with any new tool, a settling-in period is needed for trainees and trainers to become familiar with it. It was important to meet with trainees and find out their opinions on the MSF tool and issues with its use. These are easy to remedy and we plan to remove the MSF tool scoring system and have comments only, to ask educational supervisors to collate the scores and to state that it is not expected for trainees to use the tool where ward rounds do not occur in their subspecialty. It has become embedded in our practice, so we ask trainees to complete one every 6 months alongside the Royal College of Paediatrics and Child Health (RCPCH) requirements where the minimum number of workplace-based assessments is 12, but 20 are recommended.11

Just over one-third of trainees in the field testing were in year 4 (ST4) of training. This may be due to the fact that they had more opportunities to lead ward rounds compared with their more senior counterparts (ST6 and above). Senior trainees will be in a subspecialty rather than general paediatrics and therefore have more outpatient commitments. Just over half the trainees were in ST3 middle grade posts and ST4, and this may have played a part in the fact that scores did not change over time as these are the most junior middle grade doctors who have another 3–5 years before achieving completion of training.
This MSF tool is useful for demonstrating achievement of the competencies of the clinical leadership competency framework and will help to remedy previous reports of only 54% of regional senior paediatric trainees having opportunities to lead ward rounds. Leadership skills assessed by the tool include communication, team working, team management, use of resources, approachability and respect for others. Other specialties such as anaesthesiology and surgery have been putting an increasing emphasis on assessing non-technical skills such as situational awareness, team working, communication, leadership and decision-making skills, and tools have been developed—namely, the Anaesthetists’ Non-Technical Skills (ANTS) and Non-Technical Skills for Surgeons (NOTSS) systems, respectively. They allow consultants to observe and give feedback on non-technical performance of trainees in the intraoperative environment.

The finding that consultants in the group gave significantly lower scores is in agreement with the results reported by Archer et al. Significantly lower self-assessment scores than assessors’ scores for the MSF tool contrasts with the findings of Wall et al with the MSF tool Team Assessment of Behaviour (TAB), but agrees with the findings of paediatric trainees in the Sheffield Peer Review Assessment Tool (SPRAT) study. However, Davis et al found weak or no associations between physicians’ self-rated assessments and external assessments. It may be that paediatric trainees often underrate themselves because they set high targets and expect for themselves. Positive comments from the team can, however, boost confidence and enhance performance. There was quite a marked difference between what trainees wrote and assessors’ comments for the same ward round.

Studies relating to the Mini-Clinical Evaluation Exercise (Mini-CEX) and SPRAT tools show the ability of these assessment tools to discriminate between different levels of training based on performance. Our study did not show any statistically significant difference in performance between junior and senior trainees. This may be due to assessors being more stringent with senior trainees than with junior ones. A quantitative comparison of performance by score over 18 months of MSF use showed no significant difference, but qualitative analysis of free text comments showed that both the trainee and assessors felt that trainee skills were improving. In hindsight, the rating scale going from borderline to good may have attributed to the lack of score differences.

There are several potential benefits of using this MSF tool. These include focusing more on senior trainees leading ward rounds and being supervised by consultants. It is vital that trainees are encouraged and supported to lead ward rounds, particularly in light of the findings by Qureshi and Swamy that trainees had significantly lower scores than other assessors regarding ward round leadership ability. Qualitative comments were more helpful about trainees’ performance than use of a scoring system.

Main messages
- A MSF tool on ward round leadership skills was practicable and reliable when piloted and field tested among paediatric senior trainees.
- This MSF tool looks at preparation and organisation, communication, teaching and enthusiasm, team working and punctuality.
- Senior doctors (consultants, staff grade and associate specialist doctors and senior trainees) gave significantly lower scores than other assessors regarding ward round leadership ability.
- Qualitative comments were more helpful about trainees’ performance than use of a scoring system.

Current research questions
- Will removing the scoring scale from this ward round MSF tool facilitate formative comments to be provided by assessors?
- Is this MSF tool applicable to other specialties?
- Will use of the tool improve patient care and, if so, how can we assess this?

Key references

Contributors HMG helped with the initial conceptualisation of the study, was instrumental in the pilot and field testing, held the focus group and analysed the data from it, contributed to database design and data interpretation and took a major role in writing and reviewing the initial paper and wrote this current paper. IL, with help from co-authors, designed the study, collected data, analysed qualitative data from the MSF, drafted the original single paper and reviewed the final submission of this paper. DW helped with the initial conceptualisation and design of the study, statistical analyses of the data, contributed to the writing of the paper and reviewed the final submission of the paper. TB helped with the initial conceptualisation of the study, design of the database, data interpretation, pilot and field testing and contributed to the writing and reviewing of this work as well as the final submission of this paper.

Competing interests None declared.

Ethics approval Research Ethics Committee (REC) review was not formally obtained. NHS Research Ethics Service guidance was used (http://www.hra.nhs.uk/wp-content/uploads/2013/08/NRES_OPs_v5.1_2012.03.14.pdf) which states that research not requiring review by a REC (paragraph 1.90 page 52) includes research involving health service staff who are recruited by virtue of their professional role.

Provenance and peer review Not commissioned; externally peer reviewed.

REFERENCES


A multisource feedback tool to assess ward round leadership skills of senior paediatric trainees: (2) Testing reliability and practicability

Helen M Goodyear, Indumathy Lakshminarayana, David Wall and Taruna Bindal

Postgrad Med J published online May 4, 2015