Clinical skills in junior medical officers: a comparison of self-reported confidence and observed competence

Les Barnsley,[1] Patricia M Lyon,[1] Susan J Ralston,[2] Emily J Hibbert,[1] Ilona Cunningham,[2] Fiona C Gordon[2] & Michael J Field[1]

BACKGROUND The intern year is a key time for the acquisition of clinical skills, both procedural and cognitive. We have previously described self-reported confidence and experience for a number of clinical skills, finding high levels of confidence among Australian junior doctors. This has never been correlated with an objective measure of competence.

AIMS AND HYPOTHESIS We aimed to determine the relationship between self-reported confidence and observed competence for a number of routine, procedural clinical skills.

METHODS A group of 30 junior medical officers in their first postgraduate year (PGY1) was studied. All subjects completed a questionnaire concerning their confidence and experience in the performance of clinical skills. A competency-based assessment instrument concerning 7 common, practical, clinical skills was developed, piloted and refined. All 30 PGY1s then completed an assessment using this instrument. Comparisons were then made between the PGY1s' self-reported levels of confidence and tutors' assessments of their competence.

RESULTS A broad range of competence levels was revealed by the clinical skills assessments. There was no correlation between the PGY1s' self-ratings of confidence and their measured competencies.

CONCLUSIONS Junior medical officers in PGY1 demonstrate a broad range of competence levels for several common, practical, clinical skills, with some performing at an inadequate level. There is no relationship between their self-reported level of confidence and their formally assessed performance. This observation raises important caveats about the use of self-assessment in this group.

KEYWORDS education, medical/*standards; clinical competence/*standards; medical staff, hospitals/*standards; educational measurement.

Medical Education 2004; 38: 358–367
doi:10.1046/j.1365-2923.2004.01773.x

INTRODUCTION

The first postgraduate year following medical school (PGY1) is critical for the acquisition of competence and confidence in a number of important skills. Moreover, these skills must be developed in the setting of a service environment in which they are rapidly applied to real patients. The means and extent to which these skills are acquired has received relatively scant attention in the research literature. This is of more than academic interest. Good clinical practice dictates that junior medical officers be skilled and careful in the performance of invasive procedures, yet, for the purpose of term progress, global assessment of procedural skills often suffices, rather than formal and specific assessment, with feedback, of individual skills. Furthermore, in a recent study, over 90% of PGY1s reported that they were not prepared for the skills needed as an intern.[1] Quality management principles suggest that careful evaluation of these skills and the means by which they are acquired should be the subject of ongoing study.

[1]Department of Medical Education, University of Sydney, New South Wales, Australia
[2]Postgraduate Medical Council of New South Wales, Sydney, New South Wales, Australia

Correspondence: Professor Michael Field, Department of Medical Education, Edward Ford Building, A27, University of Sydney, New South Wales 2006, Australia. Tel: 00 61 2 9926 7097; Fax: 00 61 2 9438 2170; E-mail: michaelf@med.usyd.edu.au

Key learning points

Junior medical officers in their first postgraduate year demonstrate a broad range of abilities in common, practical clinical skills when formally measured using competency-based assessments

There is a poor relationship between junior medical officers' own assessment of their skills and their formally measured competence.

It is, therefore, imperative to ensure that junior medical officers are appropriately trained and skilled in the performance of their usual duties. We have previously reported that junior medical officers consider themselves reasonably experienced and are confident in a number of skill areas by the end of their first postgraduate year, and report high levels of confidence for virtually all clinical skills by the end of their second postgraduate year.[2] On the face of it, this is reassuring. The self-assessment of skills has been advanced as a cornerstone of adult education, particularly for medical practitioners.[3] It is based on the expectation that by this stage in their development, adult learners (i.e. the junior medical officers) will have learned to accurately assess and remedy gaps in their knowledge and skills. However, there is evidence that self-assessment of clinical skills may be inaccurate, particularly in those who perform poorly.[4] Moreover, qualitative studies have shown that self-evaluation is a complex process which 'can never be objective'.[5] We therefore sought to evaluate the extent to which self-reported confidence in clinical skills reflects observed, measured competence. At a practical level, if self-assessment were found to be a valid and reliable proxy for the formal, detailed assessment of clinical skills, then this would make the measurement of these issues relatively straightforward. From an educational perspective, if junior doctors were found to be able to reliably self-assess, then this principle would be strengthened and could be applied to other skills.

The setting for the research reported here is a large teaching hospital within the Sydney metropolitan area. Medical graduates undertake a balanced rotation of clinical attachments, giving them a mix of experiences in different clinical teams. We sought to study the relationship between self-reported confidence and observed competence for a number of routine procedural skills that are encountered during PGY1. Our aim was to determine whether the level of self-reported confidence matched observed competence for a number of clinical skills. Stated formally, the null hypothesis was that the distribution of difference scores between self-reported confidence and tutors' assessment of competence on a range of procedural skills would be symmetrical at about zero.

METHODS

The study consisted of 3 elements.

Questionnaire

First, we systematically applied a previously validated questionnaire[2] to a cohort of PGY1s. The questionnaire covered a number of procedural and cognitive skill areas and consisted of 2 elements. The first concerned assessment of experience in a particular procedure or clinical skill. This was rated on a 5-point Likert scale anchored to the number of times that the doctor had performed the procedure with and without supervision. The second component involved a 4-point Likert scale designed to measure the doctor's self-reported confidence in performing that particular skill. The 4 descriptors for each of the points on the latter scale were: 1 = not yet confident to do unsupervised; 2 = fairly confident to perform the procedure without supervision; 3 = confident to perform the procedure without supervision, and 4 = confident to teach the procedure (representing the highest level of confidence).

We have previously demonstrated a significant correlation between reported confidence and experience for all skill areas surveyed.[2] We then selected 7 procedural skills on the basis of 3 parameters. Firstly, they were common basic tasks required of junior medical officers; secondly the responses in our previous study had shown significant variability, and, finally, the selected skills could be feasibly assessed using simulations. All the selected skills were included in the local postgraduate medical council's list of required clinical skills.

Development of assessment tools

To determine competence in the selected skill areas we used a competency-based, objective structured clinical examination (OSCE) model. Each of the 7

Table 1

Skills used in OSCE assessments

Venepuncture
Intravenous cannulation
Basic life support (CPR)
ECG – recording
Insertion of male indwelling catheter
Nasogastric tube insertion
Blood cultures

skills listed in Table 1 was broken down into lists of discrete steps derived from existing hospital protocols. These lists then underwent a series of validation iterations. The first stage was to have the existing protocols reduced to key structural elements. This was performed by 1 author, a medical education officer and senior registered nurse, in consultation with other authors. The resulting document was then circulated for comment to the head of the appropriate clinical department (e.g. ECG competency was reviewed by the director of cardiology). The document was then reviewed by the clinical nurse consultant in the area and further modifications were made. Once all competencies had been reviewed, they were then assessed for their ability to be practically applied in an OSCE situation. This typically involved eliminating duplicate steps and attempting to reduce the number of steps to those that could be reasonably assessed.

At this stage, we performed a pilot assessment of each of the 7 competencies. The subjects were medical students and clinicians who had undergone training in each of the skills. These participants then took no further part in the study. To evaluate interrater reliability, 2 assessors were present at each station during the pilot phase. A scenario was presented and the subject was asked to perform a particular procedure arising from the scenario (e.g. take blood cultures). Each of the assessors then scored the performance of each step of the assessment as satisfactory or unsatisfactory using a scoring sheet. The assessors included both medical and nursing staff, typically one of each, briefed as to the nature of the exercise beforehand. Both written and verbal comments about the use of the scoring sheets were collected afterwards. Formal assessment of the agreement between each given task was then calculated. On this basis, further modifications were made to the assessment tools. Using the Kappa statistic, moderate to good agreement was demonstrated between assessors for each of the tools, confirming that these tools had acceptable interrater reliability.

The above procedure, whilst time consuming and somewhat cumbersome, is mandatory in the absence of a clear gold or criterion standard against which to judge the performance of any particular clinical skill. The process had the advantage of wide consultation with nursing and medical experts. An example of the final checklist used in the study is shown in Fig. 1.

Study of clinical skills

The final phase of the study involved applying the developed clinical skills assessment tools to a sample of PGY1s. This was performed over several weeks at the beginning of the last third of the clinical year. We chose this particular period because we would have expected most of the junior medical officers to have completed a number of terms and to have developed a broad and fairly stable skills base. Prior to the OSCE assessment, all available PGY1s completed the questionnaire reporting their confidence and experience in clinical skills. The checklists were not made available to the PGY1s before the OSCE. We were seeking to assess their 'native' performance, rather than recollection of a particular list. The list was, however, developed from existing hospital policies, which were the bases for the skills training the interns had received earlier in the year.

Ten minutes were allowed for each station. Live volunteers were used for the performance of ECGs and clinical skills; manikins and prosthetic body parts (e.g. a venepuncture arm or resuscitation doll) were used for other stations. The assessors were volunteers from the nursing and medical staff within the hospital. All had held senior positions and had been involved in clinical skills education. A standardised scenario was used to introduce the PGY1 to each assessment.

The original intention had been to perform these assessments over 1 or 2 nights, but it became clear that this was not practical given the varying shifts and peripheral clinical attachments of the PGY1s, as well as the considerable time involved in the process. The assessments were typically run in the early evening following a day's work, but additional sessions had to be run to accommodate PGY1s on night shifts.

	ECG			
	Assessor **Intern** **DOB**		**S**	**U**
1	Washes hands			
2	Gathers equipment			
3	Introduces self to patient			
4	Identifies patient using ID bracelet			
5	Explains procedure to patient			
6	Gains verbal consent			
7	Ensures patient has privacy (draws curtains)			
8	Ensures machine is set up and plugged in			
9	Ensures patient details have been entered into the machine			
10	Exposes patient's limbs and chest area (puts blanket over lower body)			
11	Is able to discuss removal of chest hair if necessary			
12	Uses appropriate gel, suction caps or stickers			
13	Places limb leads at equal level			
14	Positions V1 lead correctly - right sternal border, 4th intercostal space			
15	Positions V2 lead correctly - left sternal border, 4th intercostal space			
16	Positions V4 correctly - fifth intercostal space, mid-clavicular line			
17	Positions V3 correctly - midway between V2 and V4			
18	Positions V5 lead correctly - same level as V4, anterior axillary line			
19	Positions V6 lead correctly - same level as V4, mid-axillary line			
20	Instructs patient to remain still			
21	Waits for appropriate reading			
22	Verbalises how they would correct for excessive interference			
23	Prints ECG			
24	Checks time, date and patient details correct on printout			
25	Removes gel or dots with minimal discomfort to patient			
26	Covers patient, makes him or her comfortable			
27	Documents procedure in patient's notes			
28	Removes all rubbish and machinery			

Figure 1 ECG skill assessment sheet.

In the analysis, we had to compare the questionnaire results (which were scored on a Likert scale) with the outcomes of the competency assessments (scored as satisfactory or not satisfactory for a series of steps), a process which is necessarily empirical. We sought to reduce the results of the skills assessments to 4 categories that matched the Likert scale descriptors for confidence. The process was performed by investigators who were blind to the identity of the PGY1 and their confidence questionnaire results, and comprised a senior academic clinician, a research educationalist and a medical education officer with a background in nursing skill assessment and training.

Using consensus and prior to reviewing the actual scores achieved, a raw percentage score (percentage of correctly performed steps) that represented clear mastery of each skill was selected. We defined this as correct performance of the skill with no omissions or errors that were likely to affect the patient's comfort or welfare or compromise occupational health and safety issues. This was matched against a confidence

score of 4, defined as 'confident to teach the procedure'. Similarly, a percentage level below which it was felt there would be a consistent and high likelihood of compromise in any of the above areas was matched against a confidence score of 1, defined as 'not yet confident to do unsupervised'. Bands of the intervening scores were selected to reflect competence matching the intermediate confidence levels.

Raw percentage scores were then calculated for each PGY1's assessment and their results were allocated to the 4 percentage bands.

Finally, we reviewed each assessment sheet for each PGY1 in detail to ensure that the classification of each result was appropriate to the above descriptors. Although we did not numerically weight core components of each skill, we sought to ensure that classification based on the raw percentage score was not due to, say, correctly performing all non-core components of the skill but failing in the primary aim (e.g. not obtaining blood through venepuncture). On this basis, some PGY1s were reclassified.

The data were analysed using SPSS (Version 10.0). Descriptive statistics were produced and Wilcoxon's matched pairs signed rank test was used to test the null hypothesis that the distribution of difference scores was symmetric about zero.

RESULTS

Of 38 available PGY1s, 30 (79%) completed the questionnaires and OSCE assessments. They comprised 14 women and 16 men. The assessments were completed during the third quarter of the clinical year. The 8 PGY1s who did not attend the OSCE sessions failed to do so on account of annual leave, sickness, pressure of ward duties, secondment to peripheral hospitals or reluctance to participate for unstated reasons.

The combined results of the self-reported confidence and assessed competence for each of the clinical skills are shown in Figs 2–8.

These show that, for many skills, the self-reported confidence scores are high, with the exception of cardiopulmonary resuscitation (CPR) and nasogastric tube insertion. However, for all skills, the scores obtained from the OSCE process were lower than the self-reported confidence scores, particularly for skills that might be considered easier, such as venepuncture, intravenous cannulation and blood culture collection. Application of Wilcoxon's signed rank test confirms a true difference between the self-reported confidence scores and the assessed competence scores (Table 2). The null hypothesis that the distribution of difference scores between PGY1s' self-reported ratings of confidence and tutors' ratings of competence would be symmetrical at about zero is rejected.

CONCLUSIONS

We were originally concerned that this project would be compromised by there being no variation observed in the assessed skills because all the PGY1s would perform them well. However, we were surprised to find that significant numbers of PGY1s omitted key steps or performed specific components wrongly under direct observation in a simulated, clinical skills laboratory situation. We sought patterns in the types of errors made, and noted that most were made in the areas of sterile or aseptic technique, documentation and familiarity with equipment. It is axiomatic that these compromise important and potentially vital issues in health care delivery.

Before accepting these results at face value, it is important to consider whether there are any biases or measurement errors in the instruments or their application.

There is no gold standard for clinical skills assessment. Therefore, the validity of any instrument must be determined by means other than comparison to a standard. In effect, we are trying to create a standard with the reported instruments. The tools have face validity, established by consultation with medical and nursing experts. However, there are possible concerns that the performance of PGY1s in a simulated situation, particularly where there are fewer cues to correct performance (e.g. no notes at the bedside) may result in a worse performance than might otherwise take place. On the other hand, it could be argued that the simulated situation, without the stress of having a live patient, may result in enhanced performance. There are some data to suggest that medical students adapt their performance to the type of assessment instrument used (i.e. global assessment versus checklist-oriented)[6]. This would have been expected to enhance their performance in the OSCE, so is unlikely to have contributed to the present findings. The issue of use of simulations as a valid technique therefore remains unresolved, but the broad acceptance of clinical simulations supports their validity.

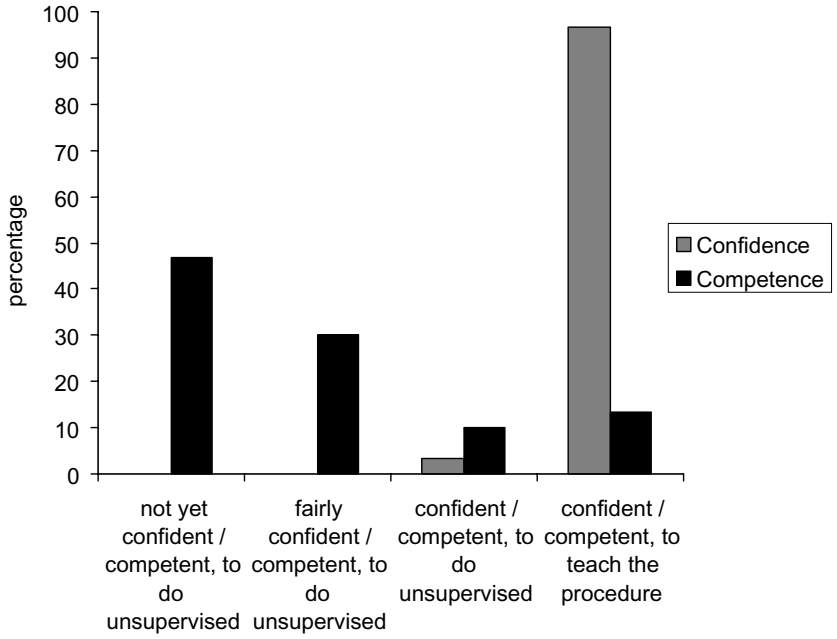

Figure 2 Venepuncture: PGY1 self-reported levels of confidence and OSCE assessment of PGY1 levels of competence.

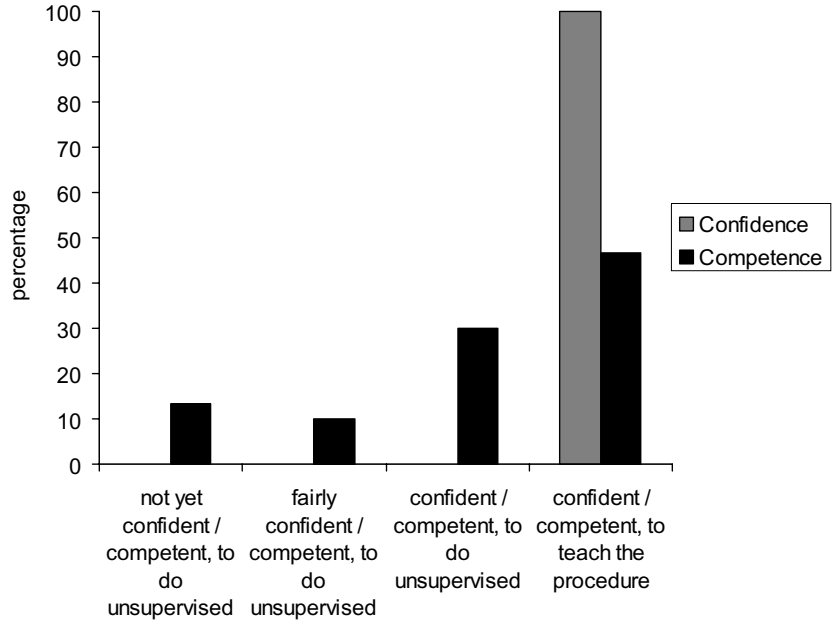

Figure 3 IV cannulation: PGY1 self-reported levels of confidence and OSCE assessment of PGY1 levels of competence.

Some of the PGY1s were unfamiliar with some of the equipment, particularly the Vacutainer system for blood and blood culture collection. Therefore, poor performance in these skills, where particular equipment is prescribed, may be falsely low. The poor performance may be a reflection of a lack of equipment-specific training rather than generic skills. However, the nature of medical procedures is that practitioners should have an intimate knowledge of the best equipment available for the task. The last few years, in particular, have seen the development of tools that help avoid needlestick injuries. It is therefore not just an issue of patient care, but occupational health and safety that requires that the most appropriate equipment be utilised. We would therefore stand by the results of assessments utilising the best current equipment.

It may be suggested that the PGY1s were 'never taught' the procedures in the detailed way in which the assessment instruments were presented. This would imply that there are less detailed, but nonetheless acceptable ways of performing the task. However, each of the steps in the competency-based

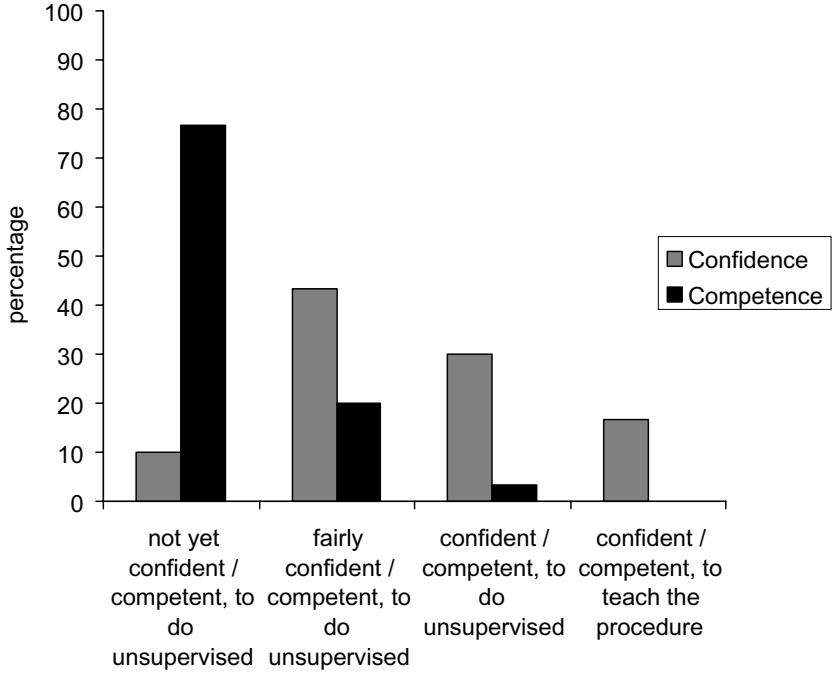

Figure 4 CPR: PGY1 self-reported levels of confidence and OSCE assessment of PGY1 levels of competence.

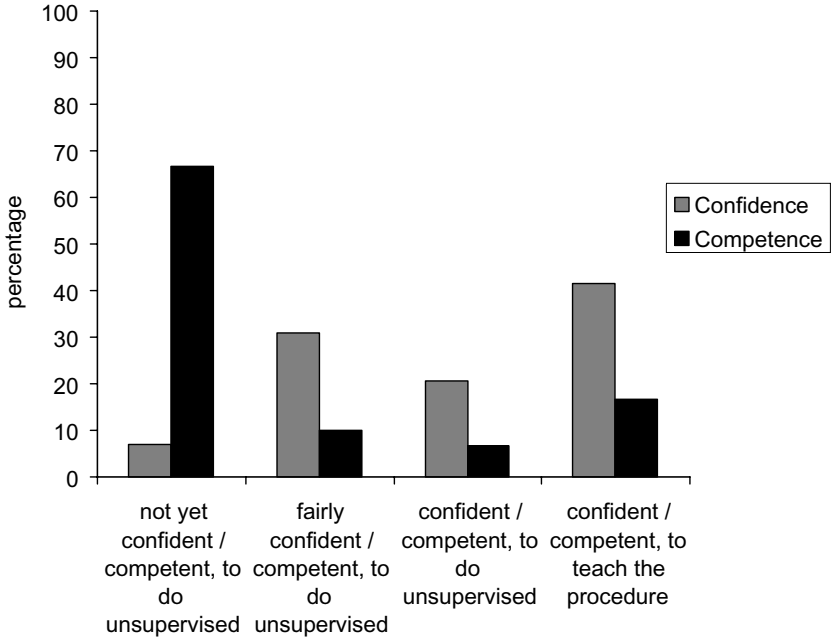

Figure 5 ECG: PGY1 self-reported levels of confidence and OSCE assessment of PGY1 levels of competence.

assessments has a specific purpose, and it is difficult to argue that leniency in any of the steps is acceptable. It is also possible that the frequent, almost mundane performance of these tasks is given little credence by the PGY1s or their supervisors. It would seem that PGY1s are expected to 'pick up' these skills in the course of their routine practice. This may be the case, and highlights the need for a more systematic education and evaluation programme for clinical skills in this group. However, this does not detract from the validity of the competency-based instrument, which is based on best practice. Further development and improvement of competency-based assessment is proposed. This will focus specifically on improving the simulations and scenarios to make them more realistic, dealing with acceptable alternative practices (especially the order in which components of a procedure are performed) and further refining the wording used at each step to facilitate more accurate marking. We will also seek to standardise the common

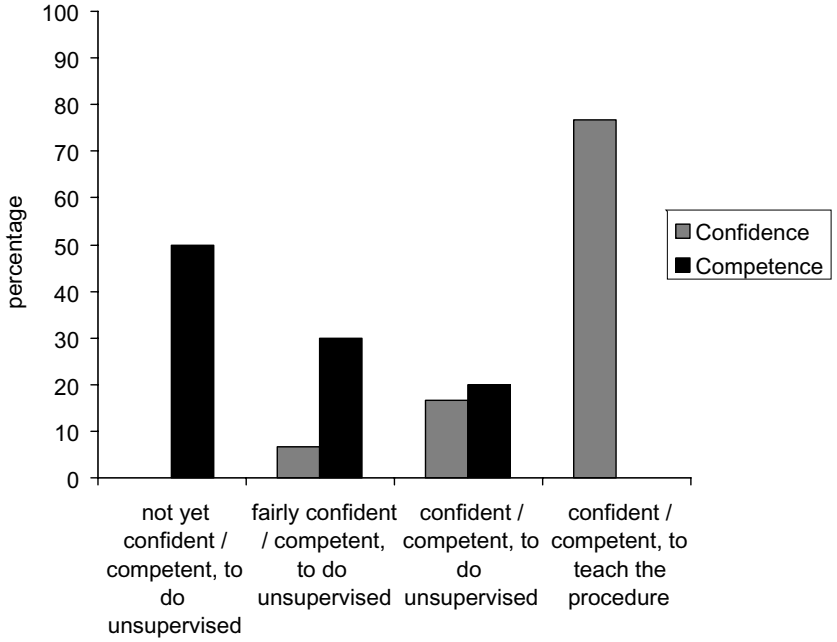

Figure 6 Bladder catheterisation (male): PGY1 self-reported levels of confidence and OSCE assessment of PGY1 levels of competence.

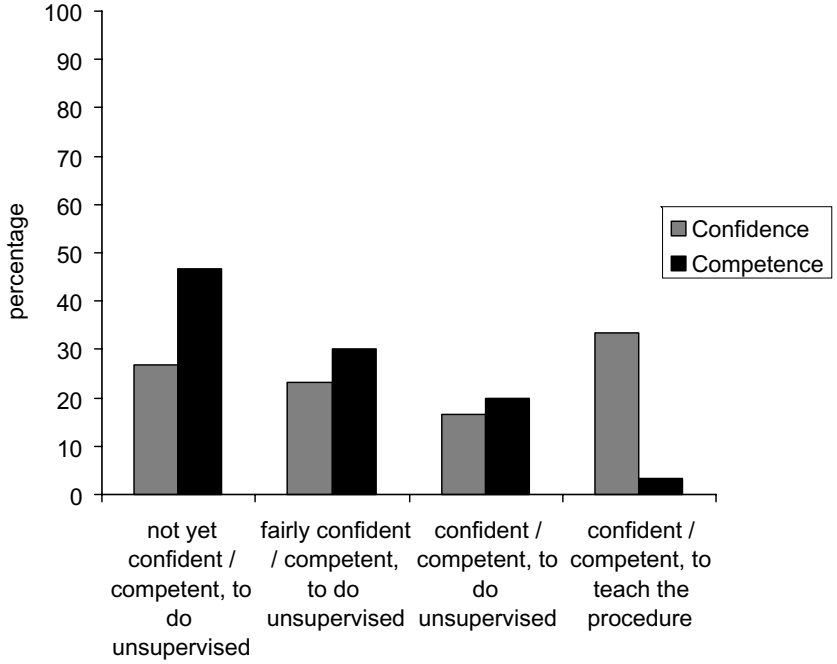

Figure 7 Nasogastric tube insertion: PGY1 self-reported levels of confidence and OSCE assessment of PGY1 levels of competence.

components of the skills so that they can be more easily taught, recalled and assessed.

Another possible source of bias or distortion of these results is the assessors. However, all assessors used in this programme were active nursing or medical clinicians, and all had been briefed on their role. The score sheets invited binary outcomes for each skill step so there seems little opportunity for subjective interpretations.

The most likely conclusion is that many of the cohort studied were performing procedural clinical skills at a level below 'best practice'. It would also seem fundamentally unlikely that the PGY1s studied here are any different to those at other hospitals, or that there is any systematic deficiency in their training that is not present in other institutions. This is not to criticise the PGY1s in any way, but gives pause for consideration of formal training in, and assessment of, such skills in the early postgraduate years.

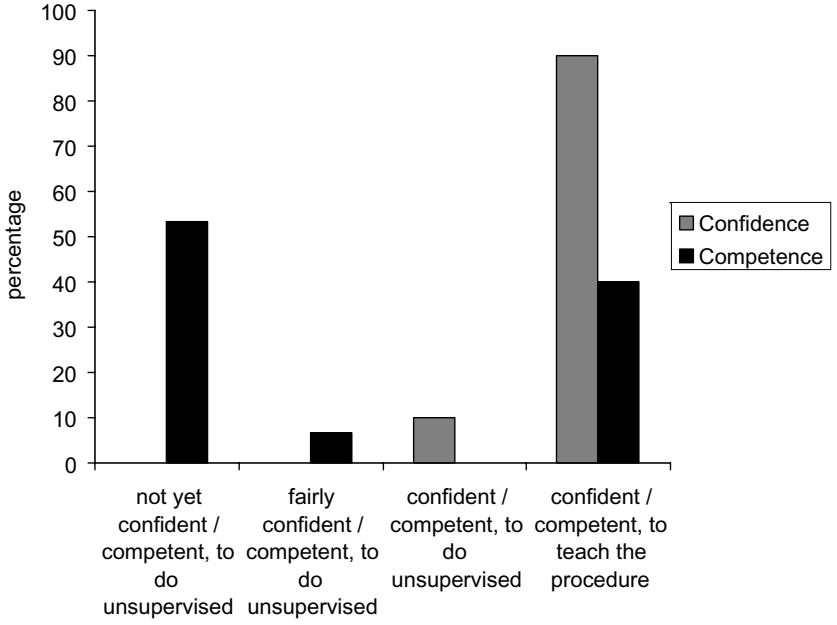

Figure 8 Blood cultures: PGY1 self-reported levels of confidence and OSCE assessment of PGY1 levels of competence.

Table 2 Differences between PGY1s' self-reported levels of confidence with selected procedural skills and levels of competence assessed by OSCE

Procedural skill	Trainee self-reported level of confidence Mean + (SD)	Tutor assessment of level of competence Mean + (SD)	Wilcoxon's Z
Venepuncture	3.97 (0.18)	1.90 (1.06)	– 4.556*
IV cannulation	4.00 (0.00)	3.10 (1.06)	– 3.598*
CPR	2.53 (0.90)	1.27 (0.52)	– 4.371*
ECG	2.97 (1.02)	1.73 (1.17)	– 3.306*
Catheterisation	3.70 (0.60)	1.70 (0.79)	– 4.769*
Nasogastric tube insertion	2.57 (1.22)	1.80 (0.89)	– 2.737†
Blood cultures	3.90 (0.31)	2.27 (1.46)	– 3.974*

* Significant at the 0.001 level (2-tailed).
† Significant at the 0.01 level (2-tailed).

We were unable to demonstrate any positive relationship between measured competence among PGY1s and self-reported ratings of their confidence in performing these skills. This raises serious concerns about the ability of PGY1s to assess their own skills accurately and indicates a need for better training in self-assessment in preparation for lifelong learning. It raises serious caveats about the use of self-assessment in this professional group, and indicates a need for reliable and valid instruments against which PGY1s can measure their performance.

The results of the present study echo those of a recent report by Fox *et al.*[7] Their study used a different skill set incorporating both cognitive and practical elements. They found that neither the PGY1s nor the consultants supervising them were able to reliably identify deficiencies in basic clinical skills prior to examination at an OSCE. Another study also found the same mismatch between confidence and competence, this time in the performance of anaesthesia simulation tasks by medical students.[8]

These findings lend weight to the concern that deficiencies in both clinical skills and self-assessment skills are widespread among junior doctors. It has been argued that medical practice is 'dominated by unreflective doing', and that the fiercely competitive nature of medical training is a barrier to revealing fears and deficiencies.[3]

The next steps are to further develop these assessment instruments and then apply them in other centres for baseline assessment, formative assessment and the evaluation of skills training programmes.

CONTRIBUTORS

This study was designed by LB, PML, SJR, EJH, IC and MJF. The OSCE stations were primarily developed by SJR and IC. The OSCEs were coordinated by SJR and FCG. All authors except PML acted as assessors at some or all of the OSCE sessions. The initial analysis of the OSCE results was performed by LB, PML and FCG. Formal Statistical analysis was performed by PML. All authors contributed to the writing up and approval of the final manuscript.

ACKNOWLEDGEMENTS

We acknowledge the help of the department heads and clinical nurse consultants who worked with us in developing the OSCE instruments. We particularly thank the PGY1s who were the subjects of this study, freely giving up their to complete the assessments.

FUNDING

This study was supported by an educational grant from the Postgraduate Medical Council of New South Wales, Australia.

REFERENCES

1 Hannon FB. A national medical education needs' assessment of interns and the development of an intern education and training programme. *Med Educ* 2000;**34**:275–84.

2 Marel GM, Lyon PM, Field MJ, Barnsley L, Hibbert E, Parise A. Clinical skills in early postgraduate medical trainees: patterns of acquisition of confidence and experience among junior doctors in a university teaching hospital. *Med Educ* 2000;**34**:1013–5.

3 Westberg J, Jason H. Fostering learners' reflection and self-assessment. *Fam Med* 1994;**26**:278–82.

4 Hodges B, Regehr G, Martin D. Difficulties in recognising one's own incompetence: novice physicians who are unskilled and unaware of it. *Acad Med (Suppl)* 2001;**76**:87–9.

5 Stewart JO, Hallorhan C, Barton JR, Singleton SJ, Harrigan P, Spencer J. Clarifying the concepts of confidence and competence to produce appropriate self-evaluation measurement scales. *Med Educ* 2000;**34**:903–9.

6 McIlroy JH, Hodges B, McNaughton N, Regehr G. The effect of candidates' perceptions of the evaluation method on reliability of checklist and global rating scores in an objective structured clinical examination. *Acad Med* 2002;**77**:725–8.

7 Fox RA, Ingham Clark CL, Scotland AD, Dacre JE. A study of pre-registration house officers' clinical skills. *Med Educ* 2000;**34**:1007–12.

8 Morgan PJ, Cleave-Hogg D. Comparison between medical students' experience, confidence and competence. *Med Educ* 2002;**36**:534–9.

Received 29 August 2002; editorial comments to authors 3 December 2002; accepted for publication 24 May 2003