IN BRIEF

- Accurate self-assessment is important for reflection, appraisal and life-long learning.
- Evidence in the literature and this study suggests poor self-assessment is common, both generally in higher education and in medicine and dentistry.
- When surgeons removed a lower third molar tooth, apparent poor self-assessment of performance was more likely to be due to the desire to convey a favourable impression than to lack of insight.
- It is possible that pressure to provide evidence of good performance will encourage surgeons to manage their image and over-score themselves.

The influence of self-deception and impression management upon self-assessment in oral surgery

A. W. Evans,¹ R. M. A. Leeson,² T. R. O. Newton John³ and A. Petrie⁴

Objective To see if poor self-assessment of surgical performance during removal of mandibular third molars is influenced by self-deception (lack of insight) and impression management (trying to convey a favourable impression).

Design A prospective study of 50 surgeons, surgically removing a lower third molar tooth.

Setting One UK dental school over a two year period.

Methods The surgeons' surgical skills were assessed (by two assessors) and self-assessed using check-list and global rating scales.

Post-operatively, surgeons completed validated deception questionnaires which measured both self-deception enhancement (lack of insight), and impression management (the tendency to deliberately convey a favourable impression).

Main outcome measures Reliability between assessors, and between assessors' and surgeons' self-assessments were calculated. Discrepancies between assessors' and surgeons' scores were correlated with surgeons' deception scores.

Results Reliability between assessors was excellent for checklist (0.96) and global rating scales (0.89) and better than the reliability between assessors and surgeons (0.51 and 0.49). There was a statistically significant correlation ($r=0.45$ $p=0.001$ checklist, $r=0.48$ $p<0.001$ global) between over/ under-rating of their surgical performance by surgeons and their impression management scores. No statistically significant correlation was found between this inaccuracy in self-assessment and surgeons' individual self-deception scores.

Conclusion The majority of surgeons scored themselves higher than their assessors did for surgical skill in removing a single mandibular third molar tooth. Impression management (the tendency to deliberately convey a favourable impression) may contribute to a surgeon's inaccurate self-reporting of performance. Lack of insight appears to be

much less important as a contributing factor. The authors speculate that pressure to provide evidence of good performance may be encouraging surgeons to manage their image and over-score themselves.

INTRODUCTION

Self-assessment, the act of judging ourselves,¹ must be perceptive and accurate if clinicians are to engage effectively in reflective appraisal, revalidation and lifelong learning. The idea of including training in self-assessment into the medical school curriculum is relatively recent and evidence suggests that self-assessment both generally and within medicine is inaccurate.² A number of reasons have been suggested for this, including not knowing what was expected and scoring effort or potential rather than actual performance.³ Doctors and more specifically surgeons have been frequently described as arrogant,⁴⁻⁶ but the present-day culture of blame may also encourage clinicians to present themselves in as favourable a light as possible.

This paper investigates whether apparent over or under marking of their own surgical skills by staff and postgraduates in the Oral and Maxillofacial Surgery Department of one UK dental hospital, is influenced by Self-Deception Enhancement (lack of insight [SDE]) or Impression Management ('faking good'[IM]).⁷

METHODS

Ethics committee approval was obtained to assess the surgical skills of staff, trainees and postgraduates whilst surgically removing a mandibular third molar tooth. In order to have a power of 80% to detect as significant, a correlation coefficient of 0.4 (between the variables of interest) at the 0.5% level of significance, approximately 46 surgeons were required. Fifty different surgeons were assessed when removing one mandibular third molar tooth. The measurement scales were based on the 'Objective Structured Assessment of Technical Skills' [OSATS]^{8,9} by one, and whenever possible, two assessors (as available). These assessors were taken from a pool of 14 assessors accustomed to supervising trainees in this procedure and experienced in these assessment methods. Surgical skill was marked by two assessors in 41 cases and by only one assessor in nine cases ie 50 surgeons were assessed in total. All surgeons completed self-assessment forms. All operations required removal of bone to extract the tooth and consecutive operations meeting the criteria were included. The OSATS method uses a checklist

¹Senior Lecturer/ Honorary Consultant; ²Specialist, Department of Oral and Maxillofacial Surgery; ³Senior lecturer and Head of Biostatistics Unit, Eastman Dental Institute for Oral Health Care Sciences, University College London; ⁴Clinical Psychologist, Pain Management and Research Centre, Royal North Shore Hospital, Sydney, NSW 2065 Australia.

*Correspondence to: Dr Ann Evans, Department of Oral and Maxillofacial Surgery, Eastman Dental Institute for Oral Health Care Sciences, University College London, 256 Grays Inn Road, London, WC1X 8LD, UK
Email: a.evans@eastman.ucl.ac.uk

Refereed paper

Received 04.12.03; Accepted 01.06.04

doi: 10.1038/sj.bdj.4812416

© British Dental Journal 2005; 198: 765–769

scale in which various stages of the operation (in this case 20) are scored as correctly or incorrectly performed and has been widely advocated for providing feedback¹⁰ and testing competency.^{11,12} The checklist used has been previously reported⁹ and includes criteria such as: appropriate design of flap; smooth reflection of flap in correct plane; correct bone removal (site and amount); tooth elevated correctly; single attempt at needle passage at correct height. This gave a score ranging from 0–20. In addition surgeons are marked using a global rating scale, anchored with descriptors, which examine eight different aspects of performance, such as respect for tissue, time and motion, handling of instruments, knowledge of procedure and use of assistants, which gave a score ranging from 8–40. This scale is claimed to be better at differentiating between levels of skill^{8,13} and for measuring ability to deal with unexpected contingencies.¹⁰ Immediately postoperatively surgeons were asked to mark themselves using the same two scales.

In order to explore possible reasons for the discrepancy between self (surgeons') and assessors' ratings, the Paulhus Deception Scale⁷ (PDS) was simultaneously administered. This is a validated 40 item questionnaire that measures an individual's tendency to give socially desirable responses on questionnaires. There are two components of this scale, Impression Management (IM) and Self-Deception Enhancement (SDE). Impression management refers to the tendency to give inflated self-descriptions by 'faking or lying' and to deliberately convey a favourable impression ('faking good') whereas self-deception enhancement indicates overconfidence and lack of insight. Surgeons were assured that all responses were completely confidential and for research purposes only. No feedback was given.

Data were analysed according to accepted statistical techniques as described in the results section. The mean difference, 95% confidence intervals, British Standards Institution Reproducibility Coefficient and a paired *t*-test were used to examine the inter-rater agreement.¹⁴ The two-sample *t*-test was used to compare subgroups. Agreement (ie reliability) between assessors was also measured using an intraclass correlation coefficient.¹⁵ This is equivalent to the reliability coefficient sometimes quoted. A value of 0 for the coefficient indicates no agreement and a value of 1 indicates perfect agreement; acceptable reliability typically requires the intraclass correlation coefficient to have a value in excess of 0.8. Correlation between various factors was evaluated using Pearson's correlation coefficient. Chi-squared tests were used to compare proportions.

RESULTS

Checklist scores (score range 0–20)

Inter-rater reliability of assessors

The intraclass correlation coefficient between the two assessors was 0.96 for the checklist scale, which suggests excellent agreement. The mean difference in score between the two assessors for the checklist was estimated as 0.24, ($n=41$, 95% CI = -0.13 to 0.62). There was no evidence of bias ($t = 1.30$, degrees of freedom = 40, $p=0.20$), which would indicate no systematic difference between the two assessors. To measure the agreement between two assessors, the estimated standard deviation of differences (S_d) provides a measure that can be used as a comparative tool. However it is more usual to calculate the British Standards Institution Reproducibility Coefficient (BSRC), a measure of agreement, which estimates the maximum likely difference between assessor one and assessor two. It is calculated by the following formula:

$$\text{British Standards Institution Reproducibility Coefficient} = 1.96 S_d$$

For the checklist scale the maximum likely difference between assessor one and assessor two was 3.9.

Fig. 1a Checklist scale

Fig. 1b Global rating scale

Difference between mean score of assessors and surgeons' self-assessment score.

The intraclass correlation coefficient between the mean of the two assessors' (or the single assessor as appropriate) and surgeons' self-assessment scores was 0.51 for the checklist scale, suggesting only moderate agreement. Overall, surgeons were inclined to mark themselves higher than their assessors and these results are summarised in Table 1. Some surgeons, who were given low scores by the assessors, over-rated themselves, sometimes by a considerable amount. The BSRC between assessors and surgeons was 6.14. For each individual surgeon 95% of the differences between assessors and surgeons will be expected to lie between the upper and lower limits of agreement and be less than 6.14 (Fig. 1a).

The difference between the mean score of the assessors and the self-assessment score of the male and female surgeons is shown in Table 2. Both male and female surgeons over-scored themselves significantly when compared to their assessors. There was a significant difference in over-scoring between males and females: difference in means (males – females) = 1.94. (95% CI = 0.26 to 3.62 $p=0.03$).

Global rating scale (Score range 8–40)

Inter-rater reliability of assessors

The intraclass correlation coefficient between the two assessors was 0.89 for the global rating scale which again suggests excellent agreement. The mean difference in score was estimated as 0.15, ($n=41$, 95% CI = -0.82 to 1.11). There was no evidence of bias, ($t=0.31$, degrees of freedom = 40, $p=0.76$).

The British Standards Institution Reproducibility Coefficient (BSRC) measures the maximum likely difference between assessor one and assessor two and was 6.10.

Table 1 Assessment of surgical skills.

	Checklist scale range (0–20)	Global rating scale range (8–40)
Mean score (95%CI)		
surgeons (n = 50)	16.68 (15.86 to 17.50)	27.76 (28.13 to 31.39)
Mean score (95%CI)		
assessors (n = 50)*	14.58 (13.45 to 15.71)	26.70 (24.84 to 28.56)
Mean difference	+ 2.10	+3.06 p
95% CI	1.26 to 2.94	1.42 to 4.70
BSRC	6.14	11.35
(paired t-test) to assess bias	t = 4.47, p<0.001	t = 3.44, p<0.001
* In 41 operations the mean score was for 2 assessors. In 9 operations the mean score was for 1 assessor only.		

Table 2 Difference between surgeons' and assessors' scores for male and female surgeons.

	Checklist score range (0 – 20)	Global rating score range (8 – 40)
Mean (surgeon – assessor). (95% CI)	2.80 (1.75 to 3.84)*	3.09 (1.04 to 5.14)*
Male surgeons (n = 32)	p < 0.001	p = 0.004
Mean (surgeon – assessor). (95% CI)	0.86 (–0.49 to 2.22)*	3.00 (–0.03 to 6.03)+
Female surgeons (n = 18)	p = 0.02	p = 0.05

* significantly different from zero.

+ not significantly different from zero.

Table 3 Distribution of IM scores

	IM		
Normal range	≤8	16/50	(32%)
Faking good – Results may be invalid	>8	34/50	(68%)
Faking good – Results probably invalid	>12	11/50	(22%)

Table 4 Distribution of SDE and PDS scores

	SDE			PDS		
Normal range	≤ 4	35/50	(70%)	≤ 11	27/50	(54%)
Above average	≥ 5	15/50	(30%)	≥ 12	23/50	(46%)
Very much above average	≥ 8	5/50	(10%)	> 16	14/50	(28%)

Differences between mean score of assessors and surgeons' self-assessment scores

The intraclass correlation coefficient between the mean of the two assessors' (or the single assessor, as appropriate) and the surgeon's self-assessment scores was 0.49 for the global rating scale, again suggesting only moderate agreement. The mean score of the surgeons (self-assessment) compared with that of their assessors is summarised in Table 1. Some of those given low scores by the assessors over-rated themselves considerably but this tendency was much less marked than for the checklist scale. A paired t-test showed evidence of bias, $p < 0.001$. The BSRC between assessors and surgeons was 11.35. For each individual surgeon 95% of the differences between assessors and surgeons will be expected to lie between the upper and lower limits of agreement (Fig. 1b).

The self-assessment score of both the male and female surgeons compared to the mean of the assessors is shown in Table 2 and suggests male surgeons over-scored themselves significantly compared to their assessors. However there was no evidence of a difference in over-rating between males and females using this scale (difference in means (males – females) = 0.09, 95% CI = –3.36 to 3.55, $p = 0.96$).

Correlation between checklist and global rating scale

The Pearson correlation coefficient between the checklist and global rating scales for all assessments was shown to be statistically significantly different from zero ($n = 141$, $r = 0.83$, $p < 0.001$).

Impression management scores (IM)

A total of 34/50 (68%) had impression management scores > 8 described by Paulhus as 'faking good – results may be invalid'. Eleven out of 50 (22%) had scores >12 described by Paulhus as 'faking good – results probably invalid' (Table 3). Five out of 50 (10%) surgeons had scores >14, described as 'very much above average' (ie scores in the highest 2.5% of the general population).

Twenty-three out of 32 (72%) male surgeons may have been 'faking good' and 7/32 (22%) were probably doing so. Eleven out of 18 (61%) female surgeons may have been 'faking good' and 4/18 (22%) were probably doing so. Three out of 32 (9%) of male surgeons were in the range described as very much above average, whereas 2/18 (11.1%) female surgeons fell into this category. There was no statistically significant difference between the proportion of males and females in either the 'maybe faking' ($p = 0.64$), 'probably faking' range ($p = 0.99$) or very much above average range ($p = 0.99$) of impression management scores. This lack of significance could be due to low power because of a relatively small sample size.

There was a significant correlation between over-rating in the checklist score and IM score (Pearson's correlation $r = 0.45$ $p = 0.001$, Fig. 2a) and also a significant correlation between over-rating in the global rating score and IM score (Pearson's correlation $r = 0.48$ $p < 0.001$, Fig. 2b). This suggests that impression management may influence how a surgeon self-assesses.

Self-deception enhancement scores

Self-deception enhancement demonstrates overconfidence in the context of poor insight. The results are summarised in Table 4. There was no significant correlation between over/ under-rating of self-assessment and self-deception enhancement scores using either the checklist or global rating scales. Checklist scale, $r = 0.34$, $p = 0.82$, global rating scale, $r = 0.20$, $p = 0.89$.

Male surgeons tended to have higher self-deception enhancement scores than females: 13/32 (41%) male and 2/18 (11%) female surgeons had scores described 'above average' for the population (> 84% of the general population), and 5/32 (16%) male and 0/17 (0%) female surgeons had scores 'very much above the average' (>97.5% of the general population).

However there was no significant difference between the proportion of males and females in either the above average ($p = 0.062$), or very much above average range ($p = 0.145$) of self-deception enhancement scores. The lack of significance could be due to low power because of a relatively small sample size.

PDS scores

The Paulhus Deception Scale [PDS] is the sum of the SDE and IM scores and thus represents the total deception score. Twenty-three out of 50 (46%) surgeons had scores >12, described as 'above average' for the general population and 14/50 (28%) had scores >16, described as 'very much above average'. This equates to a T-score >70⁷ which suggests that only 2.5% of the general population would score as highly.

There was no evidence of any statistical difference between the PDS scores of males (mean = 13.63, 95% CI 11.75 to 15.50) and females (mean = 11.17, 95% CI 8.30 to 14.04) (difference in means 2.46, 95% CI –0.75 to +5.67, $p = 0.59$). Sixteen out of 32 (50%) males and 7/18 (39%) females had above average PDS scores whereas 9/32 (28%) males and 5/18 (28%) females were in the very much above average range for the general population. There was no significant difference between the proportion of males and females in either the above average (>12) range, $p = 0.65$, or very much above average

Fig. 2a Checklist scale

Fig. 2b Global rating scale

(>16) $p = 0.10$ range of PDS scores. This lack of significance could be due to low power because of a relatively small sample size.

The correlation coefficient for over/under-rating in the checklist score and PDS score was statistically significantly different from zero ($r = 0.36$, $p = 0.001$).

The correlation coefficient for over/under-rating in the global rating score and PDS score was statistically significantly different from zero ($r = 0.37$, $p = 0.008$).

DISCUSSION

When responding to questionnaires some people give accurate insightful self-assessments; others deliberately try and manage the impression they give by describing themselves too positively (impression management). A third group will exaggerate their positive attributes but believe themselves to be honest (self-deception enhancement).⁷ We investigated whether these factors may have adversely influenced accurate self-assessment of surgical skills found in a previous study.¹⁶ We wanted to see if surgeons knowingly over-scored their performance compared with their assessors or whether they genuinely overestimated the quality of their surgery. We found that there was a good correlation between impression management and poor self-assessment, but there was no statistically significant correlation between self deception enhancement and poor self-assessment. We have assumed that assessors correctly set the 'gold standard' although we are aware that this may not necessarily be the case. However, there was good agreement between the assessors suggesting a consistent standard was being set.

It is not within the scope of this study to look at gender

differences in relation to impression management and self deception enhancement, as much larger numbers are needed. Nevertheless we have reported our findings as they seem of interest, even though not statistically significant. This is, possibly because of the small sample size in each group.

It is often claimed that self-assessment skills are poor eg level of confidence is not a predictor of outcome in either clinical or written performance assessments¹⁷ and Marteau¹⁸ described a negative correlation between students' judgement of their ability and videotaped assessment of their communication skills. When examining surgical skills, Evans *et al.* confirmed the marked tendency for poor self-assessment in weaker trainees.¹⁶ The present study confirmed this earlier work, with statistically significant over-rating of their surgical skills by surgeons compared to their assessors. Various reasons have been suggested for poor self-assessment, including not knowing what was expected, scoring effort or potential rather than actual performance and scoring high as a defence mechanism.³

We recognise that providing a numerical score rather than a qualitative evaluation of performance exerts additional pressure on assesseees. However postgraduate doctors and dentists will experience an increasing number of assessments and appraisals, which will vary from specialist examinations, to appraisal, revalidation, teaching and research assessments and verification of professional development. Although these will increasingly involve reflection and feedback, few can be protected from some numerical evaluation of performance, however much this contravenes current educational ideology. It is imperative we understand the ways individuals respond to such pressure.

Robins and Paulhus¹⁹ argue that self-enhancement is best viewed as a trait, with costs and benefits to individuals and organisations. Hays *et al.* define insight in the context of medical education and assessment as 'awareness of one's performance in the spectrum of medical practice'. It involves awareness of one's own performance, awareness of the performance of others and a capacity to reflect on both of these measures and make a judgement.²⁰ The SDE scale demonstrates overconfidence in the context of poor insight generally. High scorers, for example, claim to 'know it all' even when asked about things they could not possibly know about and can be seen as arrogant, hostile and domineering.

It claims to measure overconfidence rather than simple confidence.²¹ Although 30% of surgeons had high scores indicating unintentional overconfidence or lack of insight, there was no significant correlation between these scores and over-rating. The prevalence of overconfidence in our study was higher, on average, in men than women (although this was not statistically significant), with no female surgeons scoring in the very high range.

Impression management is an *intentional* distortion of responses to present a more positive image.²² In the IM part of the questionnaire in this study, respondents are asked to rate the degree to which they typically perform various desirable but uncommon behaviours. If they report an excess of such behaviours, they may be exaggerating and purposely trying to impress. Although unlikely, Paulhus⁷ claims a very high score may be an accurate self-assessment by an unusually saint-like person. (There is little to suggest that any of our surgeons came into this category!) It was of particular interest that a large number of surgeons (70%) appeared to score high or very high ('maybe faking', or 'probably faking') when tested for impression management, perhaps indicating that the vast majority of surgeons felt a need to present themselves in a more favourable light. The proportion of surgeons with high IM scores was similar in males and females.

Our results showed a statistically significant correlation between the over-rating of surgical skills (when self-assessing) and impression management scores, albeit with a wide scatter of the results. Others²³ have found that those claiming a high degree of ability ('high self-monitors') can use impression management

tactics more effectively than 'low self-monitors'. In particular, 'high self-monitors' appear to be more adept at using ingratiation and self-promotion to achieve favourable images among their colleagues. The tendency to 'fake good' was consistently related to agreeableness, emotional stability, and conscientiousness across a variety of job-types and organisations.²⁴ In an OSCE-type assessment McIlroy *et al.*²⁵ found evidence that medical students adapted their performance to the way they were told they would be marked. An individual's ability to adapt to the global scale depended on their expertise in a particular domain.

There is no doubt that some surgeons found these assessments stressful. Not only were they being assessed, but in addition their self-assessment was being assessed. This may have affected their performance and the pressure they felt they were under may have enhanced their need to impress or fake good. Nevertheless, these assessments were for research purposes only; if they were part of a more formal assessment or accreditation examination, the pressure to impress would surely have increased even further. The effect of stress on performance and the use of self-assessment scores for evaluation therefore need to be taken into account in any assessment of surgical skills. Even in less quantitative and more reflective evaluations we need to be aware of the possibility of faking or plagiarism in order to present a favourable impression.

In the past, surgeons have been accused of over-confidence, arrogance and lack of insight. Although 30% of the surgeons in this study showed lack of insight – that is to say they scored high or very high for self-deception enhancement – we could not find evidence to suggest this affected their opinion of their surgical performance. Seventy per cent of surgeons had impression management scores suggesting that they may have been deliberately trying to give a favourable impression. These IM scores correlated significantly with the inability to assess their own surgical skills.

This investigation did not attempt to explain why clinicians should choose to 'fake good': it merely establishes a prevalence within the group studied. It is, however, worth speculating briefly on possible causes. The tendency to 'fake good' or 'spin' has most frequently been regarded as the preserve of politicians. However in recent years the 'blame culture' has grown and competition for training and substantive posts increased; it is perhaps not surprising that clinicians feel under pressure to present themselves in the best possible light. *If this is true*, then it is not that poor surgeons are deluding themselves that they are good. It suggests that poor surgeons may know they are poor, but feel obliged to cover this up. We need to ask if we are training doctors to 'fake good' in our growing emphasis on clinical governance?

We wish to thank the staff, trainees and postgraduates who helped with this study. A. W. Evans is supported by The Health Foundation. This work was

undertaken by A. W. Evans, R. M. A. Leeson, T. R. O. Newton John and A. Petrie with University College London Hospitals who received a proportion of its funding from the NHS Executive; the views expressed in this publication are those of the authors and are not necessarily those of the Trust or the NHS Executive.

1. Boud D. *Enhancing learning through self-assessment*. London: Kogan Page, 1995.
2. Gordon M J. A review of the validity and accuracy of self-assessments in health professions training. *Acad Med* 1991; **66**: 762-769.
3. Woolliscroft J O, TenHaken J, Smith J, Calhoun J G. Medical students' clinical self-assessments: comparisons with external measures of performance and the students' self-assessments of overall performance and effort. *Acad Med* 1993 **68**: 285-294.
4. Hoffenberg R. Medical arrogance. *Clin Med JRCPL* 2001; **1**: 339-340.
5. S. Bosley. The Guardian. *Arrogance lets babies die – now a chance to mend NHS*. Thursday 19th July 2001.
6. Posen S. The portrayal of the doctor in non-medical literature. *Aust N Z J Surg* 1996; **66**: 630-650.
7. Paulhus D L. *Paulhus Deception Scales (PDS): the balanced inventory of desirable responding-7. User's manual*. New York: Multi-Health Systems Inc, 1998.
8. Martin J A, Regehr G, Reznick R K *et al.* Objective structured assessment of surgical skill (OSATS) for surgical residents. *Br J Surg* 1997; **84**: 273-278.
9. Evans A W. Assessing competence in surgical dentistry. *Br Dent J* 2001; **190**: 343-346.
10. Evans A W, Aghabeigi B, Leeson R M A, O'Sullivan C, Eliahoo J. Assessment of surgeon competency to remove mandibular third molar teeth. *Int J Oral Maxillofac Surg* 2002; **31**: 434-438.
11. Mossey P A, Newton J P. The structured clinical operative test (SCOT) in dental competency assessment. *Br Dent J* 2001; **190**: 387-390.
12. Moorthy K, Munz Y, Sarker S K, Darzi A. Objective assessment of technical skills in surgery. *Br Med J* 2003; **327**: 1032-1037.
13. Khan M S, Bann S D, Campbell-Smith T, Darzi A, Butler P E M. Validation of surgical course assessment. *Ann R Coll Surg Engl* (suppl) 2002; **84**: 173-175.
14. Bland J M, Altman D G. Statistical methods for assessing agreement between two methods of clinical measurement. *Lancet* 1986 i, 307-310.
15. Armitage P, Berry G, Matthews J N S. *Statistical Methods in Medical Research* (4th Ed). pp 704-706. Oxford: Blackwell Science, 2002.
16. Evans A W, Aghabeigi B, Leeson R, O'Sullivan C, Eliahoo J. Are we really as good as we think we are? *Ann R Coll Surg Engl*. 2002; **84**: 54-56.
17. Morgan P J. Cleave-Hogg D. Comparison between medical students' experience, confidence and competence. *Med Educ* 2002; **36**: 534-539.
18. Marteau T M, Humphrey C, Matoon G, Kidd J, Lloyd M, Horder J. Factors influencing the communication skills of first year clinical medical students. *Med Educ* 1991; **25**: 127-134.
19. Robins R W and Paulhus D L. The character of self-enhancers: implications for organizations. In Roberts B W, Hogan R (Eds). *Personality psychology in the workplace. Decade of Behaviour*. pp 123-219. DC, USA: American Psychological Association Washington, 2001.
20. Hayes R B, Jolly B C, Caldon L J M, McCrorie P, McAvoy P A, McManus I C, Rethans J-J. Is insight important? Measuring capacity to change performance. *Med Educ* 2002; **36**: 965-971.
21. Paulhus D L. Interpersonal and intrapsychic adaptiveness of trait self-enhancement: a mixed blessing? *J Personality Soc Psychol* 1998; **74**: 1197-1208.
22. Edwards A L. *The social desirability variable in personality assessment and research*. New York: Dryden press, 1957.
23. Turnley W H, Bolino M C. Achieving desired images while avoiding undesired images: exploring the role of self-monitoring in impression management. *J Appl Psychol* 2001; **86**: 351-360.
24. Weillbaecher A D. Self-report personality measures of fake good in employment selection. Dissertation – Abstracts – International: Section-B:-The Sciences and Engineering. 2001 61(9-B): 5039.
25. McIlroy J H, Hodges B, McNaughton N, Regehr G. The effects of candidates' perceptions of the evaluation method on the reliability of checklist and global rating scores in an objective structured clinical examination. *Acad Med* 2002 **77**: 725-728.