Title
The role of instructional context in medical students' self-assessments

Authors
Relan, Anju
Wilkerson, L
Doyle, Hy Lawrence
et al.

Publication Date
2006-08-01

The role of instructional context in medical students' self-assessments

Anju Relan, PhD; LuAnn Wilkerson, Ed.D.; Hy Lawrence Doyle, Ed.D., and Gretchen Guiton*,
David Geffen School of Medicine at UCLA; *University of Colorado School of Medicine

Introduction and Background

Information explosion in Medicine and the changing nature of clinical practice has warranted the need to educate independent, self directed physicians capable of superior self-assessment skills. In a recent review, Walsh (2006) contends that clinicians are more likely to change practice through assessment of their own knowledge deficits, than when this learning is imposed upon them.

While there is wide theoretical acceptance of the critical role of self-assessment in medical students' training, research on the predictive validity of this construct has been mixed. Studies over the last two decades have shown that overall, learners are poorly calibrated in assessing their own knowledge and competence (Eva, Cunnington, Reiter, Keane, & Norman, 2004; Fitzgerald, White, & Gruppen, 2003; Langendyk, 2006). An interaction between achievement and self-assessment accuracy has been documented, with high achieving students underestimating their competence, and low achievers inflating their self-assessments. Research is beginning to reveal that the knowledge needed for competence is the same as that required for self- assessment, leading some researchers to recommend de-emphasizing self-assessment as a construct, and instead expend learning resources to improve performance for the development of self-assessment skills (Eva et al., 2004; Langendyk, 2006; Regehr, Hodges, Tiberius, & Lofchy, 1996; Reiter, Eva, Hatala, & Norman, 2002).

Some studies have challenged the notion that self-assessment is poorly calibrated among learners (Lynn, Holzer, & O'Neill, 2006; Pierre et al., 2005; Weiss, Koller, Hess, & Wasser, 2005), implying that it is not clear that solely knowledge accounts for learner' self-assessment patterns. Thus self-assessment skill may interact with factors other than knowledge such as constructs in motivation and self-regulation (Regehr & Eva, 2006). In a comprehensive study of self assessments with third year medical students, Wooliscroft et. al (1993) found poor agreement between performance and self-assessments. In explaining this phenomenon, they hypothesized that students have stable internal representations of their abilities, which may not meet the realities of their performance. The consistency of students' self-assessment ratings across time, tasks and comparison with peers suggests that indeed there may be self-representation variables at play, including self-efficacy and individual attributions of success.

In this paper we explore the role of contextual instructional variables in determining self-assessment accuracy, namely effort expended on learning (as measured by time spent), and perceptions of content difficulty. In a previous study with second year medical students, we showed that self-assessments were modestly related to immediate knowledge assessment (formative assessment), but failed in distant predictions (summative assessment), based on a single self assessment estimate after formative assessments (Relan et. al, 2006). Thus immediate feedback from formative assessments was ineffective in fostering distant predictive accuracy, and contributed only partially to calibrations of near self-assessments. This research extends previous studies by exploring instructional variables that may account for medical students' assessment patterns.

The following questions framed the study:

1. Is there a difference in self-assessment accuracy on formative versus summative assessments? Are self-assessments influenced by retrospective versus prospective predictions?
2. Do effort expended and perceptions of content difficulty influence self-assessment?

Methods

Formative assessments: Seven, weekly on-line formative assessments were delivered within the eighth "Block" of a second year, integrated, problem-oriented curriculum at a leading US medical school via ANGEL course management system (http://www.angellearning.com). These assessments are designed to focus students on important content in the eight-week course, cue students to the kind of questions that may be asked on the final exam, and help students identify gaps between their learning and course expectations. The assessments are also used by instructors to evaluate instructional validity, leading to strategic interventions if the latter is deemed unsatisfactory.

Approximately 20, timed, multiple choice questions were posed within each formative assessment, designed largely to test students' clinical knowledge and skills aligned with the Block objectives. Immediate feedback on correct or incorrect responses was revealed after students submitted responses to all questions. General, elaborate feedback on each question was provided to clarify correct and incorrect choices as well as to stimulate further study from course, text and web-based resources. Students could view their own performance online compared to the class average for all assessments.

For the purpose of this study, the last two formative assessments in the Block were used to capture student self-assessments, perceptions of content difficulty and amount of time spent on self study for formative assessments.

Summative assessment: A timed, online summative assessment consisting of items similar in format to formative assessments was administered in the ninth week of the experimental Block. Students completed this synchronously in a proctored environment. An item analysis was conducted to identify and delete poor items, leading to a total of 94 items included in the study.

Self assessment probes: Questions eliciting self assessments, perceptions of content difficulty and study time were elicited after each of the two formative assessments at the end of multiple choice questions, before students received feedback. These were: 1) How would you rate this week's material? A Likert scale from 1 (Far too difficult) to 10 (Far too easy) was used to elicit responses; 2) Please estimate your percentage score on this self assessment (open ended response); and 3) Approximately how many hours did you spend on self study for this week's material? Please indicate in whole hours (open ended). The self assessment in this case was *retrospective*.

The following probe was inserted at the end of the feedback page: Please estimate what your percentage score would be on the final exam items related to this week's material (open ended). The self-assessment elicited here was *prospective*.

Finally, the summative assessment concluded with the following questions before viewing feedback: How would you rate the level of difficulty of the material in this final exam? Responses were elicited on a Likert scale from 1 (Far too difficult) to 10 (Far too easy); 2) Please estimate your **percentage score** on this final exam (open ended); and 3) Approximately how many hours did you spend on self

study for this final exam's material after the last self assessment? Please indicate in whole hours (open ended).

Subjects: 114 out of 153 second year medical students enrolled in the Block completed all questions and were included in the study. Students were requested to complete these questions as a requirement for course evaluation.

Results

Descriptive analysis

Assumptions of normality of data were tested for robustness. Reliabilities of the two formative and summative assessments (Cronbach's alpha) were .40, .60 and .70 respectively. Table 1 shows descriptive statistics for all data captured in the study.

Table 1: Means and standard deviations of performance measures on assessments, self-assessments, perceptions of test difficulty and study time.

	Form. assessment 1	**Form. assessment 2**	**Sum. assessment**
Performance scores (percent)	72.1 (11.5)	73.8 (13.6)	83.2 (5.7)
Perception of difficulty	4.0 (1.3)	3.9 (1.4)	4.6 (1.2)
Time spent on studying	17.5 (10.7)	18.1(12.9)	64.6 (30.7)
Self assessment rating (Retro)	68.8 (11.0)	68.1 (10.7)	77.7 (6.7)
Self assessment rating (Prosp)	72.9 (12.1)	73.2 (9.3)	____

Students' retrospective self-assessments of formative and summative assessments are close to their actual performance on these assessments. After viewing the feedback on formative assessments they predict higher scores on the relevant component of the summative assessment showing some effect of formative feedback. Perceptions of difficulty of formative assessments lean towards medium to high difficulty, and the final exam is rated more difficult than the formative assessments. Self evaluations of effort expended on assessments (study time) shows the greatest fluctuation within each assessment, and appears the same for both formative assessments.

Analysis of study questions

1. Is there a difference in self-assessment accuracy on formative versus summative assessments? Are self-assessments influenced by retrospective versus prospective predictions?

All self-assessments ratings were correlated with their respective formative and summative assessment performance, as shown in Table 2.

Table 2: Pertinent orrelations between self-assessment ratings and performance on formative and summative assessments.

Self-assessment ratings	Performance: Form. assessment 1	Performance: Form. assessment 2	Performance: Sum. assessment
Form Assess 1 (Retro)	**.283****	------	-----
Form Assess 1 (Prosp)	----	-----	**.02**
Form Assess 2 (Retro)	----	**.55****	-----
Form Assess 2 (Prosp)	----	----	**.08**
Sum Assessment (Retro)	----	----	**.41****

** Correlation is significant at the 0.01 level (2-tailed).

Retrospective self-assessments for formative assessments showed greater accuracy than prospective assessments to related topics in the final exam. Means of prospective self-assessment ratings versus actual performance indicate that students were conservative in self-assessment of their performance on the final exam. This could be interpreted as a pattern of self-assessments among high achieving medical students as surmised by Wooliscroft et. al (1993)- students may be subjecting themselves to stringently high standards of achievement. The retrospective self-assessment for the final exam was modestly predictive of actual performance.

2. Do effort expended and perceptions of content difficulty influence self-assessments?

Perceptions of difficulty, estimates of study time and performance on formative and summative assessments were correlated to examine relationships among these variables. These are depicted separately for each assessment in Tables 3, 4 and 5.

Table 3: Correlations between perceptions of difficulty, effort expended (study time) and performance on summative assessments.

	Difficulty	Retro. self assessment	Estimated study time	Prosp. self-assessment	Form. Assessment performance	Sum. Component performance	Sum. assessment performance
Difficulty	1	.250(**)	-.302(**)	.326(**)	.150	.028	.065
Retro. self assessment	---	1	.108	.656(**)	.283(**)	.109	.143
Estimated study time	---	---	1	.049	.015	-.158	-.073
Prosp. self-assessment	---	---	---	1	.305(**)	.019	.062
Form. Assessment performance	---	---	---	---	1	.230(*)	.319(**)
Sum. Component performance	---	---	---	---	---	1	.640(**)

** Correlation is significant at the 0.01 level (2-tailed).* Correlation is significant at the 0.05 level (2-tailed).

Perceptions of difficulty for the first formative assessment (Table 3) show modest, but significant correlations with prospective and retrospective self-assessments. Thus easier perceptions of the topic are correlated with higher self-assessments. Difficulty is also significantly correlated in the anticipated direction with study time- the easier the perceived difficulty (higher ratings), the lower the estimated study time. However, perceived difficulty is not correlated with any performance measure. Study time does not predict performance or self-assessments.

Table 4: Correlations between perceptions of difficulty, effort expended (study time) and performance on formative assessment 2 and summative assessment.

	Difficulty	Retro. self assessment	Estimated study time	Prosp. self-assessment	Form. assessment performance	Sum. assessment performance	Sum. component score
Difficulty	1	.284(**)	-.183	.318(**)	.256(**)	.150	.124
Retro. self assessment	---	1	.174	.661(**)	.554(**)	.173	.045
Estimated study time	---	---	1	.183	.204(*)	-.040	-.043
Prosp. self-assessment	---	---	---	1	.509(**)	.188(*)	.080
Form. assessment performance	---	---	---	---	1	.294(**)	.314(**)
Sum. assessment performance	---	---	---	---	---	1	.567(**)

** Correlation is significant at the 0.01 level (2-tailed).* Correlation is significant at the 0.05 level (2-tailed).

Perceptions of difficulty for the second formative assessment (Table 4) show significant correlations with prospective and retrospective self-assessments, as well as performance on formative assessment. Difficulty and study time are uncorrelated in this assessment. Study time does not show any relationship with self-assessments, but predicts formative assessment performance.

Table 5: Correlations between perceptions of difficulty, effort expended (study time) and performance on summative assessment.

	Difficulty	Estimated study time	Retro. self-assessment	Sum. assessment performance
Difficulty	1	-.152	.407(**)	.276(**)
Estimated study time	----	1	-.005	-.120
Retro. self-assessment	---	---	1	.405(**)

** Correlation is significant at the 0.01 level (2-tailed).

Difficulty level related to summative assessment significantly correlated with self-assessments and performance. Study time was not found to be predictive of self-assessments or performance.

Discussion

In this paper, we hypothesized that medical students' self-assessments are influenced by motivational and self-regulatory mechanisms, such as perceptions of task difficulty and study time spent on assessments. We investigated interactions between retrospective and prospective self-assessments, and the role of perceived task difficulty and study time in predicting performance and self-assessments, in the context of two formative and one summative assessment.

Our initial findings based on correlational analyses support outcomes of previous research, in that correlations between self-assessments and performance, although significant, were at best low to modest. Thus there is potential for training medical students to improve self-assessment skills, adapted to address the internal representations of high achieving students. Students were more skilled at predicting prior, near performance (formative assessment) than distant, future performance (summative assessment). The predictions were all lower than final performance on all assessments, indicating that medical students tend to underestimate their knowledge. This is particularly intriguing for low-stakes formative assessments, which do not count towards a final grade.

Although students underestimated their learning showing little variability in their self-assessments, the estimated study time varied considerably more in proportion to self-assessments and performance distribution. Estimates of study time may be more liable to inflated or underreported estimates given what students include as self-study. A more focused definition may be warranted in future studies. In spite of this weakness in measurement, the disproportionate time that students spend on self-study, based on their self-assessments and performance, needs to be monitored to identify poor self-regulation. In this study, we intend to conduct further analyses based on low, medium and high achievers to identify groups who might represent poor self-regulation.

Finally, the role of perceived difficulty of task was a more reliable measure of students' self-assessments and performance. Students are able to match this perception accurately with self-assessments and performance. More research is needed to uncover the value of this variable in the study of self-assessments.

In conclusion, our study supported findings from previous research on self-assessments, but also revealed new findings on the role of instructional variables such as time spent and perception of the task. It is likely that motivational variables such as self-efficacy, anxiety, goal orientation interact with self-assessment skills in ways which will reveal how to facilitate the training of future physicians towards becoming independent, self-directed learners.

References

Eva, K. W., Cunnington, J. P., Reiter, H. I., Keane, D. R., & Norman, G. R. (2004). How can I know what I don't know? Poor self assessment in a well-defined domain. *Adv Health Sci Educ Theory Pract, 9*(3), 211-224.

Fitzgerald, J. T., White, C. B., & Gruppen, L. D. (2003). A longitudinal study of self-assessment accuracy. *Med Educ, 37*(7), 645-649.

Langendyk, V. (2006). Not knowing that they do not know: self-assessment accuracy of third-year medical students. *Med Educ, 40*(2), 173-179.

Lynn, D. J., Holzer, C., & O'Neill, P. (2006). Relationships between self-assessment skills, test performance, and demographic variables in psychiatry residents. *Adv Health Sci Educ Theory Pract, 11*(1), 51-60.

Pierre, R. B., Wierenga, A., Barton, M., Thame, K., Branday, J. M., & Christie, C. D. (2005). Student self-assessment in a paediatric objective structured clinical examination. *West Indian Med J, 54*(2), 144-148.

Regehr, G., & Eva, K. (2006). Self-assessment, Self-direction, and the Self-regulating Professional. *Clin Orthop Relat Res*.

Regehr, G., Hodges, B., Tiberius, R., & Lofchy, J. (1996). Measuring self-assessment skills: an innovative relative ranking model. *Acad Med, 71*(10 Suppl), S52-54.

Reiter, H. I., Eva, K. W., Hatala, R. M., & Norman, G. R. (2002). Self and peer assessment in tutorials: application of a relative-ranking model. *Acad Med, 77*(11), 1134-1139.

Relan, A., Guiton, G., Goldhaber, J. and Sopher, M. (2006). Second year medical students accuracy of self-assessments related to formative assessments. Paper presented at the Annual convention of American Educational Research Association, Chicago, 2006.

Walsh, K. (2006). How to assess your learning needs. *J R Soc Med, 99*(1), 29-31.

Weiss, P. M., Koller, C. A., Hess, L. W., & Wasser, T. (2005). How do medical student self-assessments compare with their final clerkship grades? *Med Teach, 27*(5), 445-449.