Peer assessment of professional competence

Elaine F Dannefer,1 Lindsey C Henson,2 S Beth Bierer,3 Tana A Grady-Weliky,4 Sean Meldrum,5 Anne C Nofziger,5 Craig Barclay6 & Ronald M Epstein5

BACKGROUND Current assessment formats for medical students reliably test core knowledge and basic skills. Methods for assessing other important domains of competence, such as interpersonal skills, humanism and teamwork skills, are less well developed. This study describes the development, implementation and results of peer assessment as a measure of professional competence of medical students to be used for formative purposes.

METHODS Year 2 medical students assessed the professional competence of their peers using an online assessment instrument. Fifteen randomly selected classmates were assigned to assess each student. The responses were analysed to determine the reliability and validity of the scores and to explore relationships between peer assessments and other assessment measures.

RESULTS Factor analyses suggest a 2-dimensional conceptualisation of professional competence: 1 factor represents Work Habits and the other factor represents Interpersonal Habits, including respect and trustworthiness. The Work Habits factor had moderate, yet statistically significant correlations ranging from 0.21 to 0.53 with all other performance measures that were part of a comprehensive assessment of professional competence. Approximately 6 peer raters were needed to achieve a generalisability coefficient of 0.70.

CONCLUSIONS Our findings suggest that it is possible to introduce peer assessment for formative purposes in an undergraduate medical school programme that provides multiple opportunities to interact with and observe peers.

KEYWORDS education, medical, undergraduate/ standards; educational measurement/ *standards; professional competence/ *standards; attitude of health personnel; peer review; students, medical; reproducibility of results.

Medical Education 2005; 39: 713–722
doi:10.1111/j.1365-2929.2005.02193.x

INTRODUCTION

Medical education has increasingly focused on defining professional competencies, developing competency-based curricula and constructing viable assessment tools.1,2 Although discrete dimensions of professional competence have been delineated,3,4 strategies that simultaneously assess multiple dimensions of competence remain elusive, particularly approaches that reliably measure not only knowledge and technical skills but also interpersonal and humanistic qualities that, increasingly, are becoming explicitly recognised as core to being a doctor.2,5–10 Professional competence requires consideration of both cognitive and
interpersonal skills. Thus, the development of assessment methods that measure these characteristics becomes a central challenge for medical education.

The promising work on peer ratings for assessing cognitive and interpersonal aspects of clinical performance has generated renewed interest in its use for medical students.\(^7,8,11\) We describe the development and measurement properties of a medical student peer rating instrument that addresses hard-to-measure professional competencies. Student acceptance and perceived value of the process is also reported. We begin by reviewing the existing peer assessment literature and discuss its value for assessing both cognitive and interpersonal skills.

**Overview**

**What is already known on this subject**

Previous research on assessment methods has focused primarily on medical knowledge and clinical skills, with less attention given to assessing interpersonal aspects of performance.

Recent research has shown that peer assessment is a reliable and valid method for assessing both cognitive and interpersonal aspects of a doctor’s clinical performance.

**What this study adds**

Peer assessment for medical students has been introduced using an instrument developed to assess a broad range of expected professional behaviours.

Factor analyses suggest a 2-dimensional conceptualisation of professional competence: Work Habits and Interpersonal Habits.

**Suggestions for further research**

Future research should explore the predictive value of peer assessments conducted in Year 2 of training for performance in clinical clerkships.

**Peer assessment literature**

Peer assessment in medicine has been used for a variety of purposes, with more rigorous research occurring in the development of instruments to measure the clinical performance of practising doctors. Ramsey and colleagues\(^12,15\) demonstrated that peer assessment is a reliable and valid method for assessing 2 dimensions of clinical performance:

1. a cognitive/clinical management dimension, and
2. a humanistic/psychosocial dimension.

Peer assessment has also been adopted by some residency programmes, although evidence of its value in graduate medical education is limited.\(^14,15\) Studies with residents have found that their peer ratings correlated positively with faculty ratings of technical skills and were lower than faculty ratings of humanistic and interpersonal skills.\(^16-18\)

Peer assessment for medical students was introduced in studies concerned with finding reliable measures of student performance. Schumacher’s\(^19\) comparison of Year 4 students’ peer assessments, medical school grades and licensure examination scores found that all methods provided measures of a factor labelled ‘general medical knowledge’. Only peer ratings and, to a limited extent, Year 4 grades measured a ‘skill in patient relationships’ factor. Linn and colleagues\(^20\) developed a scale for peer ratings that also yielded 2 factors, which they called ‘knowledge’ and ‘relationships’. Korman and Stubblefield\(^21\) found that peer ratings of graduating medical students’ ‘perceived interpersonal value’ were correlated with residency programme ratings of internship performance.

Arnold and colleagues\(^22\) reported the use of peer assessments as 1 component of clinical clerkship grades. They found that peers provided reliable assessments, although they were global in nature and did not generate unique information or offer predictive values that differed from faculty measures of student performance.

Asch and colleagues\(^23\) described using peer assessment as part of a formal grading process for Year 3 students that included sharing peer feedback with students. They reported that students viewed peer feedback as more meaningful than feedback received from faculty and concluded that peer assessment contributed to helping students develop learning agendas and reinforced the importance of professional attitudes and behaviours. In another study, Rudy and colleagues\(^24\) reported the use of peer assessment in a Year 1 clinical skills interviewing course.
in which 95% of the students provided both positive and negative written feedback to peers. They suggested that the constructive comments and high response rates might be attributable to feedback training for students, the usefulness of feedback, the anonymity of responses and the formative use of feedback.24

The literature suggests that peer assessment can provide a measure of difficult-to-assess interpersonal skills at all levels of medical training and that medical students are capable of giving and receiving constructive feedback. Although peer assessment may hold promise for evaluating both technical and interpersonal skills, it seems uniquely suited to the latter.19–24

Purposes of the study

The present study describes the development and implementation of a new peer assessment protocol (PAP) for undergraduate medical students. The purposes of the study were to:

1 report initial findings about the utility of a peer assessment process;
2 compare the PAP to other competency measures, and
3 determine whether the PAP provides meaningful data about student competence.

METHODS

The PAP is part of a larger comprehensive assessment (CA) designed to assess and foster professional competence in undergraduate medical students.25 The CA is based on the belief that professional competence requires the development of habits and practices of deliberation and judgement in both social interaction and cognitive–synthetic domains.2 Assessment of these reflexive qualities demands an approach that nurtures learning and self-monitoring and encourages self-reflection and motivation. To this end, the 2-week CA is a highly structured and obligatory formative assessment. It occurs 8 weeks before the end of Year 2 after students have completed most of the basic sciences courses and a 15-month ambulatory clerkship. At the end of the CA, each student receives a report that includes feedback on all assessment components. The CA includes 8 20-minute standardised patient (SP) encounters. The SPs assess students’ communication skills using a rating scale and performance checklists to assess students’ medical history taking and physical examination skills. Each SP encounter is followed by a 30-minute, post-encounter probe (short answer and essays) designed to assess the student’s application of basic science knowledge to the encounter just completed. Additional components include computer-based multiple-choice exercises and a mock board examination (MCQ test similar to USMLE I).

Peer assessment protocol

The PAP is comprised of 14 Likert items (Table 1). An ‘unable to answer’ category is also included, as contact with a randomly selected peer is sometimes insufficient to evaluate a particular behaviour. Items were selected based on those behaviours most likely to be consistently observed by peers in the medical school environment (e.g. being prepared and responsible, etc.). An additional item, intended as a global rating, asks if the student would refer a family member or friend to the peer being assessed. The PAP includes an area for narrative comments on the strengths and weaknesses of the peer being assessed.

The PAP also includes an introduction (Table 2) describing the purpose and use of peer ratings. Students are assured that their responses are confidential and that only they will see a copy of their own peer assessment report. Students are, however, expected to discuss the report with their advisory dean. The PAP also includes questions designed to determine the rater’s degree of contact and relationship with the peer being assessed. The instrument was piloted with Year 1 medical students. The final PAP was converted to a web-based form where all students were given password access to complete the peer questionnaire securely and confidentially.

Data collection

During the CA, each student was assessed by 15–16 randomly selected peers, stratified by gender. Each peer assessor must have been a co-member of at least 1 problem-based learning group or small group. Prior to completion of this exercise, students attended an interactive workshop that reviewed principles of constructive feedback. Students completed the online peer assessment form during the first week of the CA in March 2002. A 97% response rate was obtained without taking any follow-up measures. A computer-generated peer assessment report provided each student with his or her individual mean score for each item, the item-level frequencies across the 5-point Likert scale, and written feedback on strengths and weaknesses. Class means and standard deviations for each item were also provided.
Each student's assessments were averaged across raters to form aggregated scores for each student. Exploratory factor analysis was conducted with aggregated student scores using the common model for factor extraction with promax rotation. Eigenvalues > 1 and a scree test plot were used to determine the number of factors. The structure matrix, pattern matrix, and explained factor variance were examined to name extracted factors and to assess factor stability. Cronbach’s α was computed for each scale to assess internal consistency for the sample.
Generalisability studies (G-studies) and Decision studies (D-studies) were performed for each scale to partition systematic and error sources of variability and to determine the number of raters needed for reliable peer assessments. A raters-nested-within-students (r : s) crossed with items (i) design was selected where students served as the object of measurement. Peer raters, students and items were viewed as random facets. A sample of 26 students with 2 peer raters each (52 different peer raters) was randomly selected for the G-study. D-studies were then conducted to estimate generalisability coefficients – similar to reliability indices – for scale scores when using the number of peer raters to assess the impact on score integrity.

A total of 97 students participated, which generated a total of 1519 peer assessments. In addition to ratings on the 15 items, the forms contained 1300 written comments. Of these, 18 of the 1300 comments (1.5%) were judged inappropriate (e.g. based on hearsay, personal rather than behavioural, negative comments etc.) and were deleted by a faculty committee. Finally, at the conclusion of the CA, 96 students completed an online evaluation and participated in focus groups designed to elicit student perceptions about the PAP and other components of the CA.

### RESULTS

#### Item distributions

Of the 1519 peer assessments submitted, 718 (47%) had complete data across all items, 502 (33%) lacked data for 1 or 2 items, and 299 (19%) lacked data for 3 or more items. Inspection of item frequencies...
revealed that 4 items were prone to missing data: asks for feedback (40%), works independently (15%), admits mistakes (12%), and seeks responsibility (7%). The remaining items had less than 5% missing data. A chi-squared test examining missing data patterns (complete data, missing 1–2 items, missing ≥3 items) by degree of peer relationship (friend/acquaintance) indicated that students tended to omit responses to questionnaire items when they did not classify peers as friends ($\chi^2 = 40.96, \text{d.f.} = 2, P < 0.01$). To reduce the impact of these non-random response patterns, ‘asks for feedback’ was removed as a poorly performing item, and only complete data on the other items were used for subsequent analyses (992 peer assessment ratings). The item ‘dresses appropriately’ was also removed due to its high, negatively skewed distribution. Thus, individual assessments were based on the remaining 12 items, where students had an average of 10.2 peer assessments (median = 10, SD = 2.4, range = 3–16). Refer to Table 3 for students’ mean ratings on each item.

**Factor structure**

An exploratory factor analysis of aggregated peer ratings was conducted to check for the hypothesised 2 dimensions of professional and interpersonal behaviour. For diagnostic purposes, the Kaiser–Meyer–Olkin (KMO) measure of sampling adequacy and Bartlett’s test of sphericity ($P < 0.01$) were conducted. Results showed that KMO exceeded 0.80, indicating high correlations between pairs of variables, and Bartlett’s test was statistically significant, indicating that the correlation matrix was not an identity matrix. Based on these results, factor analysis was deemed appropriate for the sample.\textsuperscript{27}

The common factor model was used for extracting factors because we did not assume that items on the

<table>
<thead>
<tr>
<th>Specific student behaviours*</th>
<th>Mean</th>
<th>Median</th>
<th>SD</th>
</tr>
</thead>
<tbody>
<tr>
<td>1 Consistently well-prepared</td>
<td>4.1</td>
<td>4.3</td>
<td>0.47</td>
</tr>
<tr>
<td>2 Identifies and solves problems</td>
<td>4.1</td>
<td>4.2</td>
<td>0.41</td>
</tr>
<tr>
<td>3 Clearly explains reasoning processes</td>
<td>4.1</td>
<td>4.3</td>
<td>0.45</td>
</tr>
<tr>
<td>4 Demonstrates respect</td>
<td>4.1</td>
<td>4.3</td>
<td>0.46</td>
</tr>
<tr>
<td>5 Seeks to understand others</td>
<td>4.1</td>
<td>4.2</td>
<td>0.45</td>
</tr>
<tr>
<td>6 Takes initiative</td>
<td>3.8</td>
<td>3.8</td>
<td>0.49</td>
</tr>
<tr>
<td>7 Shares information with others</td>
<td>4.1</td>
<td>4.2</td>
<td>0.42</td>
</tr>
<tr>
<td>8 Seeks responsibility</td>
<td>4.2</td>
<td>4.3</td>
<td>0.45</td>
</tr>
<tr>
<td>9 Asks for feedback</td>
<td>4.8</td>
<td>4.0</td>
<td>0.42</td>
</tr>
<tr>
<td>10 Trustworthy</td>
<td>4.4</td>
<td>4.5</td>
<td>0.36</td>
</tr>
<tr>
<td>11 Admits mistakes</td>
<td>4.4</td>
<td>4.5</td>
<td>0.33</td>
</tr>
<tr>
<td>12 Dresses appropriately</td>
<td>4.6</td>
<td>4.7</td>
<td>0.32</td>
</tr>
<tr>
<td>13 Behaves appropriately</td>
<td>4.3</td>
<td>4.4</td>
<td>0.38</td>
</tr>
<tr>
<td>14 Thinks and works independently</td>
<td>4.3</td>
<td>4.4</td>
<td>0.38</td>
</tr>
</tbody>
</table>

**Amount of contact with peer†**

| Times in same PBL group | 2.2  | 2.3    | 0.21|
| Times in same ICM group | 2.2  | 2.2    | 0.30|
| Times in same lab | 1.7  | 1.9    | 0.34|

**Relationship with peer**

| Degree of relationship with peer‡ | 2.3  | 2.3    | 0.33|
| Quality of interactions with peer§ | 4.3  | 4.4    | 0.37|
| How well know peer* | 2.8  | 2.9    | 0.42|
| Would refer family member to peer* | 4.1  | 4.2    | 0.50|

* Items coded as 1 low/unsatisfactory to 5 high/exceptional.
† Items coded as 1 0 times to 4 ≥3 times.
‡ Items coded as 1 minimal interaction to 4 a close friend.
§ Items coded as 1 problematic to 5 positive.
* Items coded as 1 not very well to 5 very well.
PAP were measured with perfect reliability.\(^{28}\) As indicated in Table 4, the Work Habits factor received high factor weights and correlations for items representing cognitive skills and study habits. The second factor, Interpersonal Habits, received high factor weights and correlations for skills/humanistic qualities commonly associated with dimensions of professionalism. One item, ‘behaves appropriately’, loaded highly on both factors. The results of the factor analysis were used to form 2 scales by using weights to estimate factor scores, which were moderately correlated \((r = 0.46)\).

**Generalisability and decision studies**

G-studies were conducted for each factor to obtain variance estimates for use in subsequent D-studies to determine the number of peer raters needed for reliable PAP scores. There was less variability in the relative ranking of students for the Interpersonal Habits factor (14%) as compared to the Work Habits factor (20%). Peer raters contributed substantial variability for Interpersonal Habits (26%) and for Work Habits (31%). Less variability was observed among items (≤ 5%) and for interaction between students and items (≤ 2%). The residual accounted for 42% (Work Habits) and 55% (Interpersonal Habits) of the total variance.

D-studies found that approximately 6 peer raters (Fig. 1) were needed for the Work Habits and Interpersonal Habits scales in order to achieve a generalisability coefficient of 0.70, which has been judged as sufficient for formative purposes.\(^{29}\) Reliability was not significantly enhanced for the scales as the number of peer raters exceeded 15.

**Correlation analyses**

Individual student scores for the Work Habits and Interpersonal Habits scales were estimated for each student who received 6 or more peer ratings \((n = 93)\) using weights obtained through the factor analysis. Scores were correlated with peer contact measures, peer relationship measures, and other measures of student performance to assess construct validity (Table 5). Except for the mock board scores \((n = 73)\), complete data for correlation analyses

---

**Table 4 Factor weights and correlations indicating the peer assessment instrument’s factor structure \((n = 97\) students)**

<table>
<thead>
<tr>
<th>Specific item (item number)</th>
<th>Work habits</th>
<th>Interpersonal habits</th>
<th>Work habits</th>
<th>Interpersonal habits</th>
</tr>
</thead>
<tbody>
<tr>
<td>Identifies and solves problems (2)</td>
<td>0.96</td>
<td>0.93</td>
<td>0.37</td>
<td>0.37</td>
</tr>
<tr>
<td>Thinks and works independently (14)</td>
<td>0.92</td>
<td>0.91</td>
<td>0.50</td>
<td>0.50</td>
</tr>
<tr>
<td>Clearly explains reasoning processes (3)</td>
<td>0.86</td>
<td>0.87</td>
<td>0.52</td>
<td>0.52</td>
</tr>
<tr>
<td>Consistently well-prepared (1)</td>
<td>0.84</td>
<td>0.86</td>
<td>0.44</td>
<td>0.44</td>
</tr>
<tr>
<td>Seeks responsibility (8)</td>
<td>0.81</td>
<td>0.86</td>
<td>0.73</td>
<td>0.73</td>
</tr>
<tr>
<td>Takes initiative (6)</td>
<td>0.80</td>
<td>0.80</td>
<td>0.73</td>
<td>0.73</td>
</tr>
<tr>
<td>Seeks to understand others (5)</td>
<td>0.99</td>
<td>0.33</td>
<td>0.93</td>
<td>0.93</td>
</tr>
<tr>
<td>Demonstrates respect (4)</td>
<td>0.93</td>
<td>0.37</td>
<td>0.90</td>
<td>0.90</td>
</tr>
<tr>
<td>Admits mistakes (11)</td>
<td>0.90</td>
<td>0.90</td>
<td>0.85</td>
<td>0.85</td>
</tr>
<tr>
<td>Trustworthy (10)</td>
<td>0.85</td>
<td>0.34</td>
<td>0.88</td>
<td>0.88</td>
</tr>
<tr>
<td>Shares information with others (7)</td>
<td>0.71</td>
<td>0.54</td>
<td>0.81</td>
<td>0.81</td>
</tr>
<tr>
<td>Behaves appropriately (13)</td>
<td>0.37</td>
<td>0.61</td>
<td>0.65</td>
<td>0.78</td>
</tr>
<tr>
<td>Cronbach’s (z) for each factor</td>
<td>0.94</td>
<td>0.97</td>
<td>0.97</td>
<td>0.97</td>
</tr>
<tr>
<td>Variance explained for each factor</td>
<td>55.8%</td>
<td>20.6%</td>
<td>0.46</td>
<td>0.46</td>
</tr>
</tbody>
</table>

Note: Weights and correlations <0.30 are not shown. The factor pattern matrix lists weights used to estimate factor scores. These weights, referred to as ‘loadings’ in principal components factor extraction, denote the unique contribution of each variable to the factor, demonstrating the relationship of the factors to items. The factor structure matrix, on the other hand, lists correlation coefficients indicating how each item relates across factors, which provides insight about the item’s linear relationship with other factors. Both matrices should be examined to assess factor–item relationships (see Gorsuch, 1983\(^{28}\)).
The Work Habits factor correlated significantly with other performance measures providing evidence of construct validity. Interestingly, Interpersonal Habits correlated highly with peer relationship variables but not with performance measures, thereby confirming that this factor captures aspects of professional behaviour not formally assessed using traditional measures. Interestingly, ratings of interpersonal items appeared to be a function of the quality of the student's relationships with peers rather than the amount of contact in curricular settings. The item, ‘Would you refer this future physician to a family member?’ appeared to tap both domains, as evidenced by the significant item-to-factor correlations. The amount of contact with peers in formal educational settings and the number of submitted peer assessments did not seem to have a consistent, meaningful linear relationship with other variables of interest.

### Student evaluation of peer assessment process

Students \( n = 96 \) completed an online questionnaire with ratings and written comments at the end of the CA. Students reported that peers ‘identified and documented’ more strengths (35%) and weaknesses (25%) than they had considered in their own self-assessment. A total of 38% indicated that the peer assessment report identified issues that were important enough to become part of their learning plan. More than half agreed that this approach to getting feedback was ‘helpful’ (53%), while others reported mixed feelings or disagreed (22%) with this statement. Written comments, when negative, focused primarily on the perception that some peers failed to

---

**Table 5 Correlations of peer assessment factors with other variables**

<table>
<thead>
<tr>
<th>Peer assessment factors</th>
<th>1</th>
<th>2</th>
</tr>
</thead>
<tbody>
<tr>
<td>1 Work habits</td>
<td></td>
<td></td>
</tr>
<tr>
<td>2 Interpersonal habits</td>
<td>0.45*</td>
<td></td>
</tr>
</tbody>
</table>

**Amount of contact with peer**

- Times in same PBL group: -0.06
- Times in same small group: -0.10
- Times in same lab: -0.05

**Relationship with peer**

- Degree of closeness with peer: 0.19
- Quality of interactions with peer: 0.42*
- How well know peer: 0.18
- Would refer family member: 0.83*

**Performance measures/other**

- SP rating of communication: 0.27*
- Post SP encounter probes: 0.53*
- SP scores for discrete history and physical examination elements: 0.21*
- Computer exercise score: 0.38*
- Mock examination: 0.30*
- Final grade: 0.46*
- Number of peer ratings: 0.02

* \( P < 0.01 \)
† \( P < 0.05 \)
SP = standardised patient.

Note: Correlations are based on 93 student observations except for the mock examination score, which is based on 73 observations.
give constructive feedback. All students also participated in 1 of 6 focus groups in which there was consensus that some of the written feedback was problematic. Nevertheless, students said that they found the comments more helpful than the item ratings in their peer assessment reports.

**DISCUSSION**

This study suggests that peers can distinguish between 2 elements of competence: Work Habits and Interpersonal Habits. The single item, ‘refer family member’, appears to be a good proxy for both factors, suggesting that peers take 2 factors into account when rating their peers as future doctors. Students rated low by their peers on either factor were unlikely to receive high ratings on the ‘refer family member’ item.

As can be seen in Table 5, the Work Habits factor had statistically significant correlations with other performance measures of knowledge and clinical skills, suggesting that the Work Habits factor measured domains captured in more traditional assessment methods used in the CA. The Interpersonal Habits factor, however, did not correlate with these performance measures. Instead, it correlated highly with peer relationship variables. Amount of contact with peers did not appear to be related to either of these factors. These findings provide criterion-related evidence that the PAP measures 2 constructs of professional competence.

One limitation of this study is that we report results from only 1 class and 1 year at 1 medical school. This will be partially addressed as analyses of succeeding cohorts are undertaken. In addition, results are currently being correlated with other outcome measures, such as clerkship grades, which will provide additional information about the validity of peer assessment.

We view the peer assessment exercise as a professional development intervention that makes it explicit that professional competence depends on the interconnected importance of characteristics related to work habits and interpersonal skills. We emphasise that professional competence is an irreducible synthesis of cognitive and interpersonal habits. This position guided the development of the instrument. By placing the exercise during an assessment of professional competence, we hoped to reinforce this view of professional competence early in training.

The value of any assessment method must be measured by its usefulness to future learning and its power to promote professional development. We are encouraged that more than a third of our students found the results of the peer assessment meaningful enough to incorporate feedback into their learning plans and that more than half had no reservations about its positive value. Peer assessment can be a risky undertaking. Some immature students may make inappropriate and hurtful comments, and others may have difficulty interpreting and making use of their feedback. Therefore, it is essential to cultivate an atmosphere of trust, to provide explicit rules of confidentiality, and to hold small group or individual debriefing sessions. Student input should also be sought as peer assessments are designed.30 It would be naïve to expect that any single assessment of professional competence will have a meaningful impact on the entire medical school class. However, we have demonstrated that peer assessment can be a process that fosters reflection about personal and professional qualities for a substantial proportion of students.

**Contributors:** EFD participated in the development of the project, including the instrument and methods of implementation, prepared the draft manuscript and incorporated the contributions of the co-authors. LCH participated in the development of the project, including the instrument and methods of implementation. SBB contributed to statistical analysis, the Results section and general editing. TAG-W participated in the development of the project, including the instrument and methods of implementation. SM contributed to statistical analysis and editing of the Methods and Results sections. ACN contributed to implementation and the editing of the document. CB provided consultation on statistical analysis and editing of the manuscript. RME participated in the implementation and made significant contributions to the conceptualisation and presentation of the ideas presented in the paper.

**Acknowledgements:** the authors appreciate the support and advice of Dale Danner PhD at the Warner School, University of Rochester. Laura Connard provided excellent support in the development of the project and in data collection.

**Funding:** funding was provided by the Foundation for Innovations in Post-Secondary Education, grant #B16000468, and Public Health Service Grant #5 D16HP00063-02.

**Conflicts of interest:** none.

**Ethical approval:** this research was conducted under the exempt status granted by the University of Rochester Research Subjects Review Board (RSRB), which was established and operates in accordance with the requirements of the Federal Policy for the Protection of Human Subjects (45CFR46).
REFERENCES


Received 16 September 2003; editorial comments to authors 4 May 2004, 9 September 2004, 21 October 2004; accepted for publication 19 November 2004