The Relationship Between Preceptor Expectations and Student Performance on 2 Pediatric Objective Structured Clinical Examination Stations

Michael H. Malloy, MD, MS; Linda Perkowski, PhD; Michael Callaway, MS; Alice Speer, MD

Background: We designed 2 pediatric objective structured clinical examination stations, 1 anemia case associated with lead exposure and 1 failure-to-gain-weight case associated with extended breast-feeding, to evaluate third-year medical students who had studied in pediatric community preceptors’ offices as part of a 12-week multidisciplinary ambulatory clerkship rotation.

Objective: To examine the relationship between preceptor expectations and student performance on these 2 objective structured clinical examination stations.

Methods: To elicit community preceptors’ expectations of student performance, we constructed a 46-item survey replicating checklists filled out by simulated patients evaluating student performance on the objective structured clinical examination stations. The percentage agreement among preceptors for each checklist item as well as the percentage agreement between preceptor responses and student responses on each checklist item were calculated. A summary score of preceptor responses across all checklist items and a summary score for student responses across all checklist items on each station were calculated. The correlation coefficients between preceptor and student summary scores were then examined.

Results: Fifty-nine preceptor surveys were mailed and 38 were returned (64% response rate). Data were usable from 37 surveys. Eighty-nine percent (33 of 37) of the preceptors agreed that a third-year clerkship student should have the knowledge to care for the patient with anemia and 92% (34 of 37) of the preceptors agreed similarly for the growth-delay case. Agreement among preceptors on individual checklist items varied widely for both cases. Fifty-seven students studied at the anemia station and 34 students studied at the growth-delay station. The mean ± SD agreement across the 26 items on the anemia case between preceptor responses and student responses was 62% ± 23% and, for the 21 items on the growth-delay case, 60% ± 17%. The mean ± SD preceptor summary score for the anemia case was 17.4 ± 3.8 (maximum, 26) and 16.0 ± 3.6 (maximum, 21) for the growth-delay case. The mean student score on the anemia case was 15.5 ± 3.7 (maximum, 26) and, for the growth-delay case, 10.0 ± 4.5 (maximum, 21). The Pearson correlation coefficient between the preceptor and student scores on the anemia case was 0.19 (P = .15), and for the growth-delay case, − 0.41 (P = .06).

Conclusions: These data suggest community preceptors agree on topic areas in which students should be clinically competent. There was, however, considerable variation in agreement among preceptors about what preceptors believe students should be able to do and how the students actually perform. The overall percentage agreement between preceptor expectations and student performance appears to be no better than chance.

Arch Pediatr Adolesc Med. 1998;152:806-811

Editor’s Note: The results of this study are depressing, if not surprising. I wonder what the results would have been if academic center–based preceptors were involved. Any bets?

Catherine D. DeAngelis, MD

From the Departments of Pediatrics (Dr Malloy), Family Medicine (Mr Callaway), and Internal Medicine (Dr Speer), University of Texas Medical Branch, Galveston, and the Division of Medical Education, University of Southern California School of Medicine, Los Angeles (Dr Perkowski).
MATERIALS AND METHODS

In June 1996, the University of Texas Medical Branch, Galveston, implemented its Multidisciplinary Ambulatory Clerkship. This is a third-year core clerkship in which medical students spend a total of 12 weeks studying in the community offices of pediatric, internal medicine, and family practice physicians. Time is split equally across the 3 disciplines. As part of the student evaluation process, an 8-station, end-of- clerkship OSCE was developed. Three of the stations were pediatric cases. One station was an interview station in which the presenting problem was anemia in an infant. The underlying cause of the anemia was exposure to a lead-contaminated environment. A second station, where students presented their interview findings to a faculty member and were asked several questions concerning the diagnosis of anemia in childhood, was paired with the first. The third station was an interview station in which the student received growth information and dietary information about an 8-month-old infant with a decrease in weight gain associated with prolonged exclusive breast-feeding.

The objectives of our study were 2-fold. First, we wanted to determine how closely community-based pediatric preceptors agree with each other in their expectations of student performance on items used in the pediatric OSCE stations. Second, we wanted to determine how closely pediatric preceptor expectations of student performance agreed with actual student performance. We define our terminology as follows: Percentage agreement among preceptors is a measure of the preceptors' expectations of student performance. Percentage agreement among the preceptors represents the proportion of preceptors responding positively to a checklist item. Percentage agreement between the preceptors and students measures the proportion of students and preceptors who responded similarly, either positively or negatively, to a checklist item. Percentage of agreement between preceptors and students was calculated as follows: \((a + d)/(a + b + c + d)\), where \(a\) indicates the positive response of both preceptors and students; \(b\), the negative response of preceptors and the positive response of students; \(c\), the positive response of preceptors and the negative response of students; and \(d\), the negative response of both preceptors and students.

We accomplished these objectives through the design of a survey instrument. The survey was used to obtain demographic and practice information of the pediatric Multidisciplinary Ambulatory Clerkship preceptors in the community. The survey instrument presented the case scenarios of the anemia case (Table 1) and the growth-delay case (Table 2). It also contained questions that were identical or very similar to the checklist items for the anemia and growth-delay stations that were completed by the standardized patients during the OSCE. For the survey, the anemia interview and initial evaluation station checklist items were combined. Fifty-nine community pediatric preceptors who had participated in the Multidisciplinary Ambulatory Clerkship received the survey instrument by mail. A follow-up mailing was sent to nonresponders.

Demographic and practice information of the community-based preceptors was summarized. The frequency of preceptor and student responses on the checklist items from the OSCE stations was determined, and we determined the percentage agreement between the preceptor response to individual items and the student response to the items. Because of the possibility of chance agreement between preceptor responses and student responses, \(k\) statistics were calculated for each checklist item to determine the percentage agreement above chance. The \(k\) statistic takes chance into account by the calculation of an expected value of agreement on the basis of chance alone, then subtracts that value from the observed percentage agreement. Values greater than 0.75 indicate excellent agreement above chance, while values less than 0.40 indicate poor agreement beyond chance. A preceptor summary score was derived for each case by summing the responses of the preceptors (ie, the preceptors' positive response of whether a student would be able to perform a task or obtain the necessary information indicated on the checklist was scored 1; a negative response was scored 0) over the individual items of each case. A similar summary score was calculated for the students. Correlation coefficients were used to examine the relationship between the preceptor and student summary scores. Preceptor summary scores were correlated directly with the summary score of the students who had studied in their office.

RESULTS

Of the 59 surveys mailed out, 38 were returned for a response rate of 64%. This response rate included the results of a second mailing to nonresponders. The majority of the preceptors were men aged 40 to 49 years, with a wide variation in the number of years at their practice site (Table 3). The number of students they had instructed varied widely; the average number of patients preceptors saw each day was 29.

The preceptors were asked 2 global questions about each case. The first question was, “Do you think it is likely that a student studying in your office would ever come in contact with a case like this?” For the anemia case, 68% (25/37) of the preceptors said yes, 19% (7/37) said no, and 13% (5/37) were uncertain. The second question was, “Do you think a third-year student...
on the completion of an ambulatory rotation should have the ability and knowledge to care for a case like this? A total of 89% (33/37) responded yes and 11% (4/37) were uncertain. For the growth-delay case, 97% (36/37) of the preceptors responded affirmatively to the first question and 92% (34/37) replied affirmatively to the second.

The percentage agreement among preceptors on the checklist items, the percentage of students responding positively to the checklist items, and the percentage agreement between preceptor responses and student responses to checklist items are presented in Table 4 and Table 5. For this analysis, preceptor responses were paired with the number of students who studied in their office. For example, if a preceptor had 2 students in his or her office over the past year who were respondents on the examination, the preceptor’s response was counted twice. As an example of the responses, the first item for the anemia case concerning inquiry about the type of milk showed that 89% (51/57) of the preceptors agreed that the students would inquire about the type of milk. Ninety-eight percent (56/57) of the students actually did inquire about the type of milk. However, the percentage agreement between the preceptor responses and the responses of students who had studied in their offices was only 88%. The k statistic for this particular item was 0.03, suggesting essentially no agreement above chance. In general, there was a great deal of variation between what preceptors thought students could do and how the students actually performed. The mean percentage agreement between preceptor responses and student responses for all 26 items was 62%, with an average k statistic of 0.04 and a range from −0.10 to 0.26.

For the growth-delay case, a similar variation between preceptor responses and student responses was noted (Table 5). For example, 77% (17/22) of preceptors agreed that the student would show the growth chart to the mother, while 95% (21/22) of the students actually did. As an illustration of the lack of agreement between preceptor responses and student responses, 91% (20/22) of preceptors agreed that the students would inform the mother about the use of acetaminophen following immunization in this case, while only 55% (12/22) of the students did so. The overall mean percentage agreement between preceptor and student responses for this case was 60% and the mean k score was 0.05.

The relationship between the summary scores for the preceptors and students on the anemia case is illustrated in Figure 1. Fifty-seven students completed this station. The maximum attainable score was 26. The mean preceptor score was 17.4, and the mean student score 15.5. The correlation coefficient between the 2 scores was 0.19 and did not attain significance. The relationship between the summary scores for the growth-delay case is illustrated in Figure 2. There are fewer points on this curve, because fewer students completed this station (n = 34). The maximum attainable score was 21, and the mean score for the preceptors was 16.0, compared with 10.0 for the students. The correlation coefficient was −0.41 and approached, but did not attain, statistical significance (P = .06).
tion. Thus, it is likely that they may demonstrate higher correlations with checklist items.

The formulation of OSCE stations or simulated patient cases seems to occur mainly through use of course evaluation blueprints or derivation from problems presented in a list of course objectives. The origin of the course objectives or blueprints varies considerably. In community-based clerkships, the course objectives or blueprints usually arise in the academic centers, not from the community faculty. Although ultimate responsibility for the course resides within the academic setting, review and validation of course objectives and participation in the evaluation process by community faculty might be an important educational process for both academic and community faculty.

As this study points out, the opportunity to determine community preceptor opinion on the content of the OSCE stations, as well as determine how accurately they could predict student performance, provides insight into what is likely taught in the community. The process may also serve to provide the community faculty with gold standard expectations for student performance. The preceptors may, in fact, be unaware of the rationale and context of the evaluation process. The survey of the community faculty may have served as a means for providing them with information about what the students' examinations covered and, thus, may have served as a faculty development tool. Although there was global agreement among the preceptors that the 2 cases used in the OSCE stations were cases that medical

<table>
<thead>
<tr>
<th>Checklist Item</th>
<th>Preceptor Response, No. (%)†</th>
<th>Student Response, No. (%)‡</th>
<th>Percentage Agreement§</th>
<th>$\kappa$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Inquiry about type of milk</td>
<td>51 (89)</td>
<td>56 (98)</td>
<td>88</td>
<td>-0.03</td>
</tr>
<tr>
<td>Inquiry about volume of milk</td>
<td>46 (81)</td>
<td>48 (84)</td>
<td>78</td>
<td>-0.09</td>
</tr>
<tr>
<td>Inquiry about multivitamin intake</td>
<td>38 (67)</td>
<td>48 (84)</td>
<td>61</td>
<td>0.00</td>
</tr>
<tr>
<td>Inquiry about iron supplementation</td>
<td>43 (75)</td>
<td>50 (88)</td>
<td>77</td>
<td>0.26</td>
</tr>
<tr>
<td>Inquiry about lead exposure</td>
<td>29 (49)</td>
<td>29 (51)</td>
<td>64</td>
<td>0.26</td>
</tr>
<tr>
<td>Inquiry about the environment in which the child plays</td>
<td>47 (82)</td>
<td>27 (47)</td>
<td>54</td>
<td>0.12</td>
</tr>
<tr>
<td>Inquiry about child's development</td>
<td>53 (93)</td>
<td>54 (95)</td>
<td>88</td>
<td>-0.06</td>
</tr>
<tr>
<td>Inquiry about vomiting or colic symptoms</td>
<td>30 (53)</td>
<td>33 (58)</td>
<td>60</td>
<td>0.19</td>
</tr>
<tr>
<td>Inquiry about appetite</td>
<td>37 (65)</td>
<td>48 (84)</td>
<td>59</td>
<td>-0.14</td>
</tr>
<tr>
<td>Inquiry about black or tarry stools</td>
<td>27 (47)</td>
<td>38 (67)</td>
<td>50</td>
<td>0.00</td>
</tr>
<tr>
<td>Recommends measurement of serum lead level</td>
<td>56 (98)</td>
<td>48 (84)</td>
<td>82</td>
<td>-0.03</td>
</tr>
</tbody>
</table>

Recommends the following tests: serum lead level, smear for basophilic stippling, serum erythrocyte protoporphyrin level, and serum iron level

All 4 tests                                                                  | 5 (9)                        | 0 (0)                     | 0                     | -0.10  |
3 tests                                                                      | 3 (5)                        | 5 (9)                     | 0                     | (weighted) |
2 tests                                                                      | 30 (54)                      | 15 (27)                   | 14                    | 0.01   |
1 test                                                                       | 18 (32)                      | 19 (34)                   | 13                    |        |
0 tests                                                                      | 0 (0)                        | 17 (30)                   | 0                     |        |
Recommends parental testing                                                  | 28 (49)                      | 8 (14)                    | 51                    | 0.01   |
Recommends follow-up                                                         | 56 (98)                      | 51 (89)                   | 88                    | -0.03  |

Differential diagnosis for anemia would include iron deficiency, gastrointestinal blood loss, lead exposure, or hereditary

All 4 diagnosis                                                             | 36 (64)                      | 0 (0)                     | 0                     |        |
3 diagnoses                                                                  | 9 (16)                       | 22 (39)                   | 5                     |        |
2 diagnoses                                                                  | 8 (14)                       | 32 (57)                   | 11                    | ...    |
1 diagnosis                                                                  | 2 (4)                        | 2 (4)                     | 0                     |        |
None                                                                        | 1 (2)                        | 0 (0)                     | 0                     |        |
Recommends best way for diagnosing iron-deficiency anemia                   | 34 (61)                      | 41 (73)                   | 59                    | 0.09   |

is by treating with iron supplementation and observing for response

For response to iron-supplementation therapy suggests following hematocrit and reticulocyte levels and checking reticulocyte levels in 5-7 d

All 3                                                                       | 20 (36)                      | 3 (5)                     | 4                     |        |
2                                                                           | 19 (34)                      | 19 (34)                   | 9                     |        |
1                                                                           | 17 (30)                      | 30 (54)                   | 16                    |        |
None                                                                        | 0 (0)                        | 4 (7)                     | 0                     |        |

*Checklist items for the anemia case include questions concerning case history and initial examination.
†Preceptor response represents the proportion of preceptors agreeing that the student would be able to respond correctly to the checklist item.
‡Student response represents the proportion of students responding correctly to the checklist item.
§Percentage agreement between preceptors and students measures the proportion of preceptors and students who responded similarly. That is, preceptors could have indicated that students would respond appropriately to the checklist item and the student did, or the preceptors indicated that they thought the student would not respond in the intended fashion and the student did, in fact, not respond in the intended fashion. The overall mean (SD) for percentage agreement and $\kappa$ were 62 (23) and 0.04 (0.12), respectively.
Ellipses indicate no statistic could be calculated because of an excessive number of empty cells.
students should have knowledge about and be able to manage (89% agreement for the anemia case and 92% agreement for the growth-delay case), there was considerable variation in agreement among the preceptors as to the particular knowledge the students would have to obtain to effectively diagnose the conditions and manage these cases. This variability among experts of clinical standards has been documented previously.17

Only by extensive training and by limiting the number of experts do reliability measures appear to improve markedly.13

In summary, we have surveyed a group of community preceptors to obtain their opinion about how students who have studied in their offices would perform on 2 OSCE stations. The results suggest disparity between what the preceptors think the students can do and what the students actually did. These results imply that further education of the community faculty about course objectives is necessary and that community faculty may serve as a valuable resource to help validate academic faculty perceptions of what information may be important to teach.

Accepted for publication May 4, 1998.

This work was supported in part by a Robert Wood Johnson Generalist Physician Initiative Grant.
REFERENCES