Description of an examination for the objective assessment of history-taking ability

R. E. NOWOTNY† and D. I. GROVE‡

†Department of Postgraduate Medical Education, Sir Charles Gairdner Hospital and
‡Department of Medicine, University of Western Australia, Australia

Summary
A novel examination of the ability of final-year medical students to take a targeted history is described. One hundred and nine students were examined in 1 day. Each student interviewed three patients, each with a different problem. One of these persons was a real patient, while the other two were simulated patients. Patients were able to give consistent histories and the use of simulated patients enabled a high degree of standardization to be achieved. The examination provided a useful degree of discrimination among students; the marks scored followed a normal distribution with a mean value of 22.2 and a range from 16 to 29 out of a maximum of 30 marks. We believe that this examination is valid, reliable and practical.

Key words: Educational measurement; methods; medical history taking; education, medical undergraduate; clinical competence; Australia

Introduction
The taking of a history is the major component in the diagnosis and management of many patients (Hampton et al., 1975). In common with many British medical schools, the final-year clinical examinations in medicine at the University of Western Australia for many years consisted of a combination of a 'long' case and several 'short' cases (Stokes, 1973). A few years ago, the long case examination was abandoned on a number of grounds: (1) it involved a large organizational burden; (2) it was considered to provide little information in addition to that given by the short case examination; and (3) the large variability in both patients and examiners led to considerable loss of objectivity.

Since the short case examination is concerned solely with the elicitation of physical signs, concern was expressed at the absence of any attempt to evaluate a student’s ability to take a history. Furthermore, it was felt that this may have been diverting the attention of students from acquiring history-taking skills in favour of physical examination and laboratory investigation.

Consequently, an attempt has been made to devise a new examination of history-taking ability. Recently, objective, structured clinical examinations in which a variety of clinical tasks were performed in sequence by all students have been described (Harden et al., 1975; Harden & Gleeson, 1979; Newble, Hoare & Elmslie, 1981). We modified this technique to test history-taking skills alone. In order to achieve this aim, we chose to use both real and simulated patients.

Method
A short list of common and important diseases in which the diagnosis is made predominantly from the history and which could be simulated, was considered. Patients with these conditions were sought from both in-patient and out-patient sources. The criteria required for acceptance of these persons included...
that they be willing to participate, sufficiently intelligent and articulate to give a consistent history and to have no obvious physical evidence of disease. Six patients were found who satisfied these criteria. They were suffering from the following conditions: infectious mononucleosis, angina, chronic obstructive airways disease, grand mal epilepsy, bronchial asthma and diabetes mellitus complicated by hypoglycaemic episodes. There were four women and two men whose ages ranged 18–69 years.

A detailed history was taken from each patient. From this history the presenting complaint and items of information considered to be particularly relevant to establishment of a diagnosis were determined. A shortened history was then taken on several further occasions to check consistency. Using this material a questionnaire was formulated. This consisted of nine statements to which the students could respond with true/false/don’t know; a tenth question asked what the most likely diagnosis was considered to be. An example of such a questionnaire is given in the Appendix. Each correct response to the true/false items scored 1 mark, while an incorrect answer was penalized 1 mark. A response of ‘don’t know’ received a score of zero. A further mark was given if an appropriate diagnosis was made. Thus the possible range of scores was from −9 to +10.

People of the same sex and similar age to the patients were sought in the general community to act as simulated patients. They did not have a medical background. Again, these persons were required to be relatively intelligent, articulate and willing to participate. Two simulated patients were matched to each real patient. Each pair of simulated patients met the real patient and observed a detailed history being taken from this patient by one of us (R.E.N.). During this interview the simulated patients were able to take notes if they wished. This was followed by general discussion of the illness and the patient’s history. After this discussion each simulated patient received a prepared summary of the real patient’s history. It was suggested to each simulated patient that they spend a little time thinking about how it might feel to be that person and have that illness, but they were asked not to attempt to obtain technical information about the condition. They were advised that any questions that they might be confronted with by the students in the examination which had not been discussed earlier, or included in the summary, should be dealt with by recourse to their own history. The details of the simulated history were learnt over the next several weeks. Within the week preceding the examination each simulated patient was interviewed in order to determine the accuracy of the learnt history.

One hundred and nine final-year medical students were examined in 1 full day. The students were informed before the examination that they would have 5 minutes to take a history which they considered relevant to the presenting complaint that they would be given. This would be followed by a 5-minute interval during which they would complete a questionnaire. They were also told that the patients they would interview might be real or simulated. Three real and six matched, simulated patients representing three illnesses were interviewed by fifty-four students in the morning and the other three real and six simulated patients were interviewed by fifty-five students in the afternoon. Thus, each patient was interviewed eighteen times. The examination was conducted in the out-patient department of the Sir Charles Gairdner Hospital with each patient being placed in a separate room. Each student interviewed three patients representing each of the three illnesses in either the morning or afternoon session. In every case, two of these patients were simulated and one was real. On entering the room the student could read the patient’s presenting complaint, e.g. ‘Blackouts’ from a card displayed on the desk. Following each 5-minute interview, the student proceeded to an adjacent room to complete the questionnaire. In this way three interviews and three questionnaires were completed in half-an-hour by each student.

At the conclusion of the examination session, each patient was asked to fill in the questionnaire themselves in order to check the reliability of the answers they were giving. If necessary, questions which were too technical to be understood by the patients were explained in lay language. Finally, students were asked to complete a separate questionnaire in which we sought their views on the acceptability of the examination and their opinion as to whether each patient they saw was real or simulated.

Results

Each question was validated by determining the number of incorrect responses recorded by the students for each patient. The number of such incorrect responses was then compared among the three patients (one real and two simulated) for each illness. In only one instance did responses given by
any patient differ markedly from those given by the other matching two patients. Subsequent inspection of the answer questionnaire completed by this patient (who was simulated) revealed that incorrect information was being given on that item. This question was deleted and scores adjusted accordingly.

The mean scores given by each group of students for each illness are indicated in Table 1. The variation between the lowest and highest mean scores was 19%; this was statistically significant ($P<0.001$, $t$ test). The mean scores for each patient are also shown in Table 1. In each of the six illnesses, no significant differences were found between real and simulated patients. The variation within each patient over time was analysed by comparing the mean student scores of the first nine interviews (before a tea break) with those of the subsequent nine (or ten) interviews. In six of the eighteen patients, the mean score after the tea break fell while the mean score rose after the interval in the other twelve patients. In no case was the change statistically significant ($t$ test).

The variation between morning and afternoon examinations was investigated. The mean total score ($\pm$ s.d.) for the fifty-four students in the morning session was 22.8 ($\pm$ 2.85) compared with 21.7 ($\pm$ 2.89) for the fifty-five students in the afternoon examination; this difference was not statistically significant. When the average scores for the interviews were analysed, however, a significantly greater value was found for the interviews conducted in the morning ($P<0.05$, $t$ test). The 162 interviews in the morning examination gave a mean score ($\pm$ s.d.) of 7.59 ($\pm$ 1.43) compared with a mean score of 7.25 ($\pm$ 1.68) for the 165 interviews in the afternoon session.

When the scores gained by each student for his three interviews were plotted, a normal distribution was seen. The values ranged 15–29 out of 30 marks with a mean value of 22.2, mode of 23 and median score of 23.

The ability of students to perceive whether a patient was real or simulated was determined by analysing their responses to the questionnaires they were given after the examinations. With respect to the six real patients, 16.5 ($\pm$13.5) % of students thought that these patients were simulated. With respect to the twelve simulated patients, 29.8 ($\pm$17.7) % of students considered that these patients were simulated. This difference was not statistically significant.

### Table 1. Student scores analysed for each illness and for individual patients

<table>
<thead>
<tr>
<th>Illness</th>
<th>Patients</th>
<th>$n$</th>
<th>Mean</th>
<th>Patients</th>
<th>$n$</th>
<th>Mean</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Real</td>
<td>54</td>
<td>7.04 ($\pm$ 1.35)</td>
<td></td>
<td>18</td>
<td>6.83 ($\pm$ 1.76)</td>
</tr>
<tr>
<td></td>
<td>Simulated (a)</td>
<td>18</td>
<td>6.94 ($\pm$ 1.16)</td>
<td>Simulated (b)</td>
<td>18</td>
<td>7.33 ($\pm$ 1.03)</td>
</tr>
<tr>
<td>2</td>
<td>Real</td>
<td>54</td>
<td>8.13 ($\pm$ 1.29)</td>
<td></td>
<td>18</td>
<td>8.17 ($\pm$ 1.38)</td>
</tr>
<tr>
<td></td>
<td>Simulated (a)</td>
<td>18</td>
<td>8.17 ($\pm$ 1.20)</td>
<td>Simulated (b)</td>
<td>18</td>
<td>8.06 ($\pm$ 1.35)</td>
</tr>
<tr>
<td>3</td>
<td>Real</td>
<td>54</td>
<td>7.61 ($\pm$ 1.50)</td>
<td></td>
<td>18</td>
<td>7.33 ($\pm$ 1.03)</td>
</tr>
<tr>
<td></td>
<td>Simulated (a)</td>
<td>18</td>
<td>7.50 ($\pm$ 2.07)</td>
<td>Simulated (b)</td>
<td>18</td>
<td>8.00 ($\pm$ 1.03)</td>
</tr>
<tr>
<td>4</td>
<td>Real</td>
<td>55</td>
<td>6.85 ($\pm$ 1.80)</td>
<td></td>
<td>18</td>
<td>6.39 ($\pm$ 1.82)</td>
</tr>
<tr>
<td></td>
<td>Simulated (a)</td>
<td>18</td>
<td>7.06 ($\pm$ 2.11)</td>
<td>Simulated (b)</td>
<td>19</td>
<td>7.11 ($\pm$ 2.21)</td>
</tr>
<tr>
<td>5</td>
<td>Real</td>
<td>55</td>
<td>6.98 ($\pm$ 1.75)</td>
<td></td>
<td>19</td>
<td>7.00 ($\pm$ 1.35)</td>
</tr>
<tr>
<td></td>
<td>Simulated (a)</td>
<td>18</td>
<td>7.22 ($\pm$ 1.52)</td>
<td>Simulated (b)</td>
<td>18</td>
<td>6.83 ($\pm$ 1.95)</td>
</tr>
<tr>
<td>6</td>
<td>Real</td>
<td>55</td>
<td>7.91 ($\pm$ 1.25)</td>
<td></td>
<td>18</td>
<td>7.94 ($\pm$ 1.16)</td>
</tr>
<tr>
<td></td>
<td>Simulated (a)</td>
<td>19</td>
<td>8.11 ($\pm$ 1.02)</td>
<td>Simulated (b)</td>
<td>18</td>
<td>7.68 ($\pm$ 1.53)</td>
</tr>
</tbody>
</table>

$P<0.001$ when compared with Illness 4, $t$ test.
Discussion

In any examination, a number of characteristics need to be considered. These include validity, reliability, practicality and usefulness.

With respect to validity, we cannot use the present study to predict the eventual performance of students as doctors, nor are we attempting to analyse the relation of this examination to other forms of testing, but we believe that the content of the examination is valid. The clinical problems selected were both common and important. Furthermore, the data which students were expected to elicit were highly relevant to clinical practice. This is in sharp contrast to the frequent necessity to use patients with rare or chronic conditions in long case examinations.

Reliability can be defined as the consistency with which a test measures a given variable. A major factor contributing to the reliability of this examination was the use of simulated patients. For our purpose, a simulated patient may be defined as a person who has been trained to reproduce accurately selected aspects of a real patient's history. Similar techniques have been used or described by a number of medical educators (Barrows & Abrahamson, 1964; Barrows, 1968; Jason & Tichov, 1971; Newble, 1977). With one minor exception, our simulated patients gave the same responses to students as did the real patients. Furthermore, students were unable to differentiate between real and simulated patients. Thus, we were able to provide each student with an almost identical examination.

Neither real nor simulated patients appeared to change their history during repeated interviews, as the scores before and after each tea break were similar. Even though the illnesses presented in the morning and afternoon examinations were different, the scores obtained by students in each session did not differ significantly. There was a slight (up to 19%) but significant variation in mean scores among the six illnesses. The effect of this variation was reduced by ensuring that each student had three interviews.

The major limitation of using only real patients is that each person would have to be interviewed an inordinate number of times by students to achieve the same degree of standardization. Since real and simulated patients were effectively the same, we were able to reduce the number of interviews per patient by two-thirds. Students interviewed each patient for only 5 minutes. This period was chosen for two reasons: (1) it was necessary logistically to enable the examination to be completed within one day, and (2) this short period approximates the time available to take a targeted history in many general practice consultations.

Nevertheless, this examination has a number of problems: (1) stringent criteria are necessary in choosing patients and simulators; (2) there may be logistical difficulties in introducing real and simulated patients to one another; (3) considerable time must be spent by the organizers in preparing for the examination; and finally (4) appropriate facilities (in this case, eighteen rooms) are necessary.

This examination enabled a differentiation of students. The spread of scores followed a normal distribution with a reasonably wide range. It is of interest that all students gained at least 50% of the marks. While this might generate some criticism, the result should not be unexpected. Final-year medical students should be able to take an adequate history in such circumstances.

In conclusion, we have described a novel examination designed to test aspects of a student’s ability to take a clinical history. We believe it has unusually high validity and reliability for a clinical examination and is quite practical. It may be a useful supplement to other forms of examination in obtaining a broad assessment of students’ clinical skills. Further studies are required to determine the relationship between the information provided by this test and that given by established examinations.

References


Appendix

The following is an example of a questionnaire completed by students after an interview.

Student's Name.......................................................... Student's Number..........................................................  
Patient No. 3                                                                                     Problem = CHEST PAIN

Please tick the correct box:

1. The patient's pain radiates down both arms
   True □    False □    Don't know □

2. The pain is often brought on by exertion
   True □    False □    Don't know □

3. The patient is unable to sleep lying flat due to shortness of breath
   True □    False □    Don't know □

4. He currently suffers with swelling of his ankles
   True □    False □    Don't know □

5. He has been a heavy smoker
   True □    False □    Don't know □

6. He has a past history of hypertension
   True □    False □    Don't know □

7. He has a family history of heart disease
   True □    False □    Don't know □

8. He is a retired farmer
   True □    False □    Don't know □

9. His current treatment includes bronchodilators as well as anginine
   True □    False [□]    Don't know □

10. What is the most likely diagnosis?..........................................................