Teaching medical ethics

Measuring the ethical sensitivity of medical students: a study at the University of Toronto

Philip C Hébert, Eric M Meslin and Earl V Dunn  University of Toronto

Authors’ abstract

An instrument to assess ‘ethical sensitivity’ has been developed. The instrument presents four clinical vignettes and the respondent is asked to list the ethical issues related to each vignette. The responses are classified, post hoc, into the domains of autonomy, beneficence and justice. This instrument was used in 1990 to assess the ethical sensitivity of students in all four medical classes at the University of Toronto. Ethical sensitivity, as measured by this instrument, is not related to age or grade-point average. Sensitivity increases between the 1st and 2nd year and then decreases throughout the rest of undergraduate medical training, such that the 4th-year students identify fewer issues than those entering medical school. Students expressing a career choice of family medicine identify more issues than their peers. Several problems with the use of the instrument and the interpretation of the data were found. Nonetheless, these findings, if reproducible, are important and their meaning needs further discussion.

Introduction

Ethics has become a common topic of discussion within the medical community. In response to various concerns, more and more universities have introduced a new or extended curriculum in ethics for medical students (1). In spite of these efforts, which take a lot of time, for both staff and students, few studies have been done to evaluate the effects of curriculum content or format. A computer search using the search terms ‘ethics’ and ‘evaluation studies’ conducted in February 1991 revealed only two articles in the MEDLINE literature that had a specific evaluation component for ethics curricula. Berseth and Durand (2) compared the attitudes toward resuscitation between paediatric residents who attended a series of seminars extending over one year with non-paediatric residents working on the same wards who did not attend. On six different scales they showed significant differences between groups on only one scale after one year. Self, Wolinsky and Baldwin (3) compared two different methods of teaching moral reasoning to medical students. They found that both methods significantly increased ‘moral reasoning’, with the case-study method being more effective than a lecture format.

A preliminary consideration of methods that might be used to evaluate the expanding ethics curriculum in the Faculty of Medicine at the University of Toronto indicated that all the reported methods were problematic. Because of the problems in assessing moral judgement we decided to attempt to assess one of the basic requirements for an ethical response to clinical problems: the ability to recognise that a moral issue exists. We have called this 'ethical sensitivity' in order to reflect the ability to discern that a clinical situation raises considerations with moral content. Using this interpretation, sensitivity does not necessarily require that a dilemma or conflict exist. Nor does it refer to the capacity to make ethically defensible clinical judgements, to resolve moral dilemmas, to analyse concepts, to come to closure, or provide a justification for action.

In 1990 we published a paper that described our preliminary attempts to develop an instrument to ascertain the ethical sensitivity of medical students (4). We recognised then that there were several problems we needed to address if any conclusions were to be validly drawn from our results. For example, we suggested that the instrument needed to be modified by 1) providing for better sampling of the groups to be evaluated; 2) controlling for potential bias by instructors who either under- or over-emphasise the instrument’s importance; 3) increasing the number of vignettes tested; 4) using a programmed series of vignettes, and 5) administering the vignettes at an earlier point in the year (4). Guided by these concerns, we refined the original instrument and, during 1990, used it to test the ethical sensitivity of a large number of medical students in all classes at the University of Toronto. Therefore, this paper has two objectives: first, to describe the instrument and some results, comparing the results from the different classes of medical students, second, to address some of the difficulties we encountered, both in the design and the analysis of the instrument, from a philosophical perspective.

Key words

Medical ethics education; evaluation; students.
Development of the instrument

In our original work we developed five vignettes and in 1989 presented, in a random fashion, one of these to medical students in the first three classes of the medical school (4). The students were asked to list the issues involved in the management of the patient presented. We then, post hoc, classified the responses into the domains of autonomy, beneficence and justice corresponding to three fundamental ethical principles described in the literature of bioethics. Since Beauchamp and Childress (5) identify a fourth principle, non-maleficence, we discuss below the implications of classifying responses into three domains, rather than four. A brief analysis of the data was conducted. Using that experience, we created a new instrument. This contained a revised version of four of the five original vignettes (see Appendix I for the full instrument). [Appendices and tables are displayed at the end of the paper, Editor]. In 1989 we presented only one vignette to each student, whereas the revised instrument presents all four vignettes to each student. Since it might be possible that there would be an ordering effect we created 24 different versions of the instrument containing all the possible sequences of the four vignettes. The students were asked to respond to all four vignettes and, as in the original instrument, to list the ethical issues pertaining to the vignette. They were not asked to make judgements but only to identify the issues.

A ‘gold standard’ was created through a consensus of the five members of the working group. The number of ‘gold standard’ issues as agreed to by the panel are shown in Table 1 (see Appendix II for our full evaluation scheme). There were a total of 29 issues in the four vignettes, spread over the three domains as shown. By vignette, the number of issues varied from six to nine. Note that the fourth vignette did not have any justice issues.

This consensus was used to calculate the percentage of issues identified but was not used in the coding of the results or in setting limits to the number of issues considered. On only a few occasions students identified issues that the consensus panel had not considered and these issues were counted equally with the others (see Results). The markers simply decided in such cases that this was a valid issue that the expert panel had missed in its gold standard. In other cases, where a student wrote down two apparently different issues that were the same, the markers decided to count this as one issue.

Methods

The instrument was given to all students entering all four classes of medical students at the University of Toronto in 1990. The 4th-year class was surveyed in June 1990 and the 1st, 2nd, and 3rd-year classes were surveyed in September 1990. The instrument was given to all students on the first day of the year during student orientation. Thus, the 1st-year students received the vignettes before they had received any medical education; the 2nd-year students after they had completed their first year of medical school (largely basic sciences but it included a brief course on ethics). The 3rd-year students had some clinical exposure and the 4th-year students, while they had a little more, had yet to do their ‘clinical clerkship’ training. All the students received the vignettes in a large room with an invigilator on hand and were expected to return the forms after a brief time (15-20 minutes).

The returned forms were divided equally into three groups and the responses were coded into the three domains of autonomy, beneficence and justice by three markers (EM, PH and ED). Each of the three markers coded all the responses for the 2nd-year class to test for inter-rater reliability. All students were also asked to provide their age, sex, grade-point average, hospital of training and their probable career choice. The coded responses and the demographic data were entered into SYSTAT, a computerised package which performs most standardised statistical texts, and analysed.

Results

We received 498 completed forms. The response rate was dependent on the method used to collect the survey forms. In years 1 and 4 the staff collected the forms and we received 183 surveys from 1st-year (79.5 per cent of the class) and 214 (86 per cent) from the 4th-year. In years 2 and 3, due to the absence of usually reliable faculty, it was left to the student class-representative to collect the forms. In year 2 we received 78 responses (31 per cent) and in year 3 only 23 responses (9 per cent).

The students varied considerably in the way they responded to the vignettes. Table 2 displays the number of issues identified by vignette for all respondents. The total number varied from 0 to 17 and all vignettes showed considerable variability. Table 3 presents the number of issues by domain. Note that within the domain of beneficence, one student identified more issues than did the panel. Note also, that the gold standard for the fourth vignette did not include any issues of justice. Table 4 displays the percentage of issues identified by vignette. There is considerable variation within domain by vignette.

Table 5 displays the percentage of issues identified by each class from both the 1989 and the 1990 years. There is remarkable similarity of the percentages though the instruments are slightly different and the 1989 results were based on a convenience sample rather than a random selection of students. The percentage of issues identified increases between 1st and 2nd year and then falls throughout the next two classes so that the 4th-year class identifies the least number of issues in these vignettes.

The ethical sensitivity of these medical students was not related to their age or grade-point average. There was no relation between the ethical sensitivity of the male students and the female students in the 4th-year
class. There were complex differences between the sexes in the earlier years but we are still analysing the results in this area.

The students were asked to indicate their probable career choice. The only career-choice group large enough to compare with the whole group were those who stated their probable career choice was family/general practice. In 4th-year students 73 (34 per cent) indicated they were planning to enter family/general practice. These students identified significantly more issues than all other students in their class, both in total and for all domains, as seen in Table 6. Table 7 displays the average total number of issues identified by the three classes; years 1, 2 and 4. In year 1 the percentage of students giving family practice as their career choice was relatively small (many students did not have a definite choice) but those indicating family medicine did identify more issues though the results were not statistically significant. In years 2 and 3 the samples were somewhat small and only a few students identified family practice as a career choice and so comparisons were not done (for example only two students in year 3 identified family practice as a career choice).

Discussion
In our previous study using a convenience sample we found that medical students' ethical sensitivity increased from year 1 to year 2 and then decreased in year 3. In this study we duplicate that finding and show a further decrease in sensitivity in year 4. Since these findings are from two cross-sectional studies they may be questioned. Nonetheless, since the results are from two consecutive years there is an element of longitudinality to the findings. This supports the hypothesis that ethical sensitivity decreases after the first year and should make the findings stronger and more credible. We are not aware of other studies in the medical area that have looked at the ethical sensitivity of students.

Berseth and Durand (2) and Self et al (3) found no relation between age, sex and grade-point average and moral decision-making. We also found no relation between ethical sensitivity and either age or grade-point average. We found a complex relation between gender and ethical sensitivity and are continuing to explore this relationship. There are two general reasons why this issue is important. First, there is an emerging discussion in the literature about the process of moral development and reasoning in men and women that may bear directly on our results (6,7). This may have implications for the design of medical school curricula and the training of doctors (8,9). A second, and perhaps more pragmatic reason is that we acknowledge the potential for gender bias in this very research: the instrument was designed by men and the results were graded and interpreted by men. We do note that the vignettes include both male and female patients.

The number of responses from the 2nd and 3rd-year classes in this study was not adequate so we have tended to underplay their significance. Nonetheless, the results from these limited numbers do tend to confirm the trends from the 1st and 4th-years. In addition the results in these two middle years are similar to the findings we obtained in the 1989 study. There is no reason to suspect systematic bias from the class representatives who did not turn in the forms so that we feel these results in years 2 and 3 may well represent their classmates. They should be interpreted with caution. Indeed there are several reasons to interpret our data with caution.

First, the vignettes were developed by an expert panel, each of us representing a different field of expertise. Apart from the potential for bias that resulted from our shared gender, we share other characteristics: we are all white professionals, with advanced degrees. It is still something of an open question whether we had the requisite expertise to construct vignettes that would permit for the best extraction of ethical issues.

A second, but related reason to interpret our findings with caution lies in the instrument we have used. Competence in ethical matters requires a multi-dimensional analysis. As for other domains of clinical competence, ethics should be evaluated in three areas: knowledge, attitudes, and behaviour (10). Our use of vignettes in assessing the ethical sensitivity of medical students seems to address only a portion of the cognitive component of ethics: the ability to recognise an ethical issue. The practical advantage of using cases to do so, however, is that it seems closer to clinical reality than, say, giving students multiple-choice tests. This model also is consistent with the revival of casuistic reasoning in bioethics (11).

The unanswered empirical question is whether there is any correlation between analysing a case for its ethical issues and performing ethically in practice. The unanswered philosophical questions are both general and particular. The general question is: can one actually 'measure' an attribute such as ethical sensitivity? It does give us pause to reflect on our calculations — down to three decimal places — of this domain of clinical competency. The more particular question relates to our results; what should count as a 'good' answer to the vignette? Does the student who identifies more issues than her colleague merit a higher grade?

We surmise here that acting appropriately in practice requires some ability to recognise an ethical issue. If one fails to recognise an ethical issue, one is less apt to be sensitive to the nuances of its resolution. The relation between knowledge and practice is far from simple as one can be quite knowledgeable about a moral issue in the abstract and yet fail utterly to handle it appropriately in practice. This is because moral behaviour requires, as well, pertinent skills, attitudes and virtues. In future studies we hope to look at this vexing domain of assessing ethical sensitivity in vivo.

One more note of caution about our analysis must be
sounded, one that demonstrates both the limitations of and opportunities for collaborative work by doctors and philosophers in medical ethics. We began with cases and then parsed them using three major principles of bioethics. While the use of three principles has been endorsed for apprising other matters in bioethics, it is still of interest to philosophers and doctors whether there are three fundamental principles in bioethics or four, the fourth being non-maleficence. Indeed, this latter principle, understood as the duty not to harm or to reduce or remove harms may be regarded as the first principle of medical ethics. A further problem with our selection of principles was in their interpretation. More than once during our assignment of responses to domains, there was disagreement amongst the markers whether a response should fall under the category of beneficence or justice. This did not affect our overall results, since we were assessing raw scores, rather than the ability to place issues into domains. We also read responses which, while fitting nicely into one of the three categories, were clearly unethical. For example, in response to the vignette involving the elderly patient at home, one student suggested that, regardless of whether or not she was competent, she should be forcibly restrained and brought into hospital!

A final problem relates to the recent criticism in the literature of these principles as a basis for ethical theory in medicine. We would agree with critics who argue that these principles are of limited use. Indeed our use of them for purposes of analysis is highly artificial. We imposed them as an easy grid for marking and analysis. We do not suppose for a moment that we have discovered something deeply telling about the evolution of medical students' sensitivity concerning the content of each principle. That is, because these principles are artificial, we do not believe we can draw any firm conclusions about, say, the evolution of students' attitudes toward autonomy or justice.

However, we do believe that the overall scores represent an important finding. They suggest that after their first year at the University of Toronto students decline in their ethical sensitivity. One reason for this could be that, at the time of this study, there was no systematic attempt throughout the years to reinforce the teachings of the 1st-year course in ethics. A required course in the 3rd-year was introduced in 1991, but none of the tested students had taken it. A curriculum renewal now underway at the University of Toronto may help address this by introducing ethical training throughout the four years of medical school. Perhaps in this way ethics will not be 'swamped' by more traditional medical school subjects.

Our study supports the findings of Sulmasy et al, who found that house officer knowledge of ethics declined with postgraduate years. Unless continuously re-affirmed, the ethical competence of medical learners—whether undergraduate or postgraduate—will tend to decline. Medical educators interested in cultivating the ethical sensitivity of medical trainees must learn how to correct for this. Certainly, it will be a challenge to find ways of imparting ethics that medical trainees will find helpful for, and not burdensome to, their clinical development. Recently, the proceedings of several conferences that have addressed this issue have been published but much work needs to be done in this area.

Conclusions

This study demonstrates that an instrument asking students to identify moral issues using vignettes does discriminate between students and gives interesting findings. The interpretation of these findings is problematic. Nonetheless, the instrument appears to be useful and we expect that further study will help to explain the changes in students' moral reasoning and will assist in the education of students to enhance their management of patients in a humane and caring manner.

Ethical sensitivity in medical students is one aspect of moral judgement that should be of interest to medical educators and to moral philosophers. In spite of several methodological and conceptual problems, these studies show a pattern that is disturbing; that is, the ethical sensitivity of medical students seems to decrease with more time in medical school. Is this a consequence of medical socialisation and is it harmful? We feel this is such a critical issue that our studies should be repeated in other settings and if they are confirmed then more thought must be given to the effects of medical education and the consequences of this education.

Acknowledgement

We acknowledge the assistance of Dr Niall Byrne and Dr Ross Reid of the University of Toronto in the early phases of this study and in the development of the instrument. Funds from the Penta Stolp/Sunnybrook Primary Care Research Trust Fund are gratefully acknowledged.

Philip C Hébert, PhD, MD, is Assistant Professor, and Earl V Dunn, MD, is Professor, in the Department of Family and Community Medicine, University of Toronto, Canada. Eric M Meslin, PhD, is Assistant Director of the Centre for Bioethics, and Assistant Professor in the Department of Philosophy, University of Toronto.

References


Appendix I: The vignettes

Students received the following scenarios in varying order and were given a half-page or so to reply to each. They were also asked for their age, sex, medical-school grade average, and probable career choice.

1. A 58-year-old male has suffered an extremely large haemorrhage in the left side of his brain and requires life support and surgical intervention to survive. He has previously discussed with you and his family his wish never to be placed on life support. However, his wife wishes to have ‘everything’ done for her husband.

2. You visit an 82-year-old female on a house call. She has been previously healthy and independent. She has been deteriorating for five weeks, she has lost weight, is mildly delirious, and has taken to her bed. You think she needs to come into hospital for further investigations of this. She adamantly refuses.

3. A patient of yours, a 24-year-old married man without any symptoms, with three children, has some blood tests as part of an insurance examination. The results reveal that he is HIV-positive (positive as a carrier of the AIDS virus).

4. Mr Stephens is a hospitalised 45-year-old widower. Tests have revealed that he has an inoperable malignant lung tumour. He is likely to die within the next six months. He has a past history of depression and is currently on anti-depressant medication. Mr Stephens’ adult children have confided to you that he cannot handle ‘bad news’ and have requested that under no circumstances should he be informed of his diagnosis. The nurses believe not only that he is capable of understanding the implications of his test results but also that he should be fully informed.

Appendix II: ‘The gold standard’ marking scheme

1. Case re: Cerebral bleed

**Autonomy:** patient’s previously expressed wishes, relevance of family wishes (proxy consent), patient may not have been against temporary life-support (3 marks).

**Beneficence:** wife’s wishes as expressing patient’s best interests, risks v benefits of intervention (viz, life-support) (2 marks).

**Justice:** patient’s age, cost concerns (2 marks).

Total: 7 marks

2. Case re: Elderly female at home

**Autonomy:** right of competent patients to make riskier choices or refuse life-saving therapy, the right to privacy or to be left alone, the need to assess this patient’s competence, look for others (family) to get proxy consent (4 marks).

**Beneficence:** duty to rescue those at risk of harm, the benefits of hospitalisation v its harms, better for patient to be treated at home (3 marks).

**Justice:** appropriate resource allocation for the elderly, unfairness to others if treated at home (burden on relatives) or in hospital (may become a ‘bed-blocker’) (2 marks).

Total: 9 marks

3. Case re: HIV+ patient

**Autonomy:** duty of confidentiality, right of privacy (2 marks).

**Beneficence:** wife’s right to know to protect her health, prevention of harm to others at large or to society (2 marks).

**Justice:** any weighing of patient’s confidences or autonomy against rights of others to be protected from harm, ‘unfair’ not to tell (2 marks).

Total: 6 marks

4. Case re: Patient with lung tumour

**Autonomy:** patient’s right to information, the professional obligation to disclose, need to assess the patient’s competence, relevance of professional norms against deception, nurses’ v family’s view re patient’s wishes (5 marks).

**Beneficence:** weigh hazards of disclosure, ‘do no harm’ (2 marks).

Total: 7 marks
Table 1  The gold standard  
(\textit{Number of issues identified by the panel})

<table>
<thead>
<tr>
<th>Vignette</th>
<th>Autonomy</th>
<th>Beneficence</th>
<th>Justice</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>3</td>
<td>2</td>
<td>2</td>
<td>7</td>
</tr>
<tr>
<td>2</td>
<td>4</td>
<td>3</td>
<td>2</td>
<td>9</td>
</tr>
<tr>
<td>3</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>6</td>
</tr>
<tr>
<td>4</td>
<td>5</td>
<td>2</td>
<td>0</td>
<td>7</td>
</tr>
<tr>
<td>Total</td>
<td>14</td>
<td>9</td>
<td>6</td>
<td>29</td>
</tr>
</tbody>
</table>

Table 2  Issues identified by vignette – all classes  
(\textit{Number})

<table>
<thead>
<tr>
<th>Vignette</th>
<th>Mean</th>
<th>Minimum</th>
<th>Maximum</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>2.292</td>
<td>0</td>
<td>6</td>
</tr>
<tr>
<td>2</td>
<td>2.084</td>
<td>0</td>
<td>6</td>
</tr>
<tr>
<td>3</td>
<td>2.055</td>
<td>0</td>
<td>5</td>
</tr>
<tr>
<td>4</td>
<td>2.087</td>
<td>0</td>
<td>5</td>
</tr>
<tr>
<td>Total</td>
<td>8.576</td>
<td>0</td>
<td>17</td>
</tr>
</tbody>
</table>

Table 3  Issues identified by domain – all classes  
(\textit{Number})

<table>
<thead>
<tr>
<th>Domain</th>
<th>Mean</th>
<th>Minimum</th>
<th>Maximum</th>
<th>Possible</th>
</tr>
</thead>
<tbody>
<tr>
<td>Autonomy</td>
<td>4.232</td>
<td>0</td>
<td>9</td>
<td>14</td>
</tr>
<tr>
<td>Beneficence</td>
<td>4.050</td>
<td>0</td>
<td>11</td>
<td>9</td>
</tr>
<tr>
<td>Justice</td>
<td>0.294</td>
<td>0</td>
<td>3</td>
<td>6</td>
</tr>
<tr>
<td>Total</td>
<td>8.576</td>
<td>0</td>
<td>17</td>
<td>29</td>
</tr>
</tbody>
</table>

Table 4  Issues identified by vignette – all classes  
(\textit{Per cent})

<table>
<thead>
<tr>
<th>Vignette</th>
<th>Autonomy</th>
<th>Beneficence</th>
<th>Justice</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>.369</td>
<td>.513</td>
<td>.080</td>
</tr>
<tr>
<td>2</td>
<td>.291</td>
<td>.304</td>
<td>.032</td>
</tr>
<tr>
<td>3</td>
<td>.374</td>
<td>.618</td>
<td>.035</td>
</tr>
<tr>
<td>4</td>
<td>.242</td>
<td>.437</td>
<td>—</td>
</tr>
<tr>
<td>Total</td>
<td>.302</td>
<td>.450</td>
<td>.049</td>
</tr>
</tbody>
</table>

Table 5  Issues identified by year: comparison of 1989 and 1990 data  
(\textit{Per cent})

<table>
<thead>
<tr>
<th>Year</th>
<th>1989</th>
<th>1990</th>
<th>P-value*</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>.30</td>
<td>.30</td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>.39</td>
<td>.37</td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>.31</td>
<td>.31</td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>—</td>
<td>.28</td>
<td></td>
</tr>
</tbody>
</table>

Table 6  Responses by career choice: 4th-year class  
(\textit{Number})

<table>
<thead>
<tr>
<th>Career Choice</th>
<th>Total</th>
<th>Autonomy</th>
<th>Beneficence</th>
<th>Justice</th>
<th>P-value*</th>
</tr>
</thead>
<tbody>
<tr>
<td>Family Medicine</td>
<td>9.000</td>
<td>4.547</td>
<td>4.120</td>
<td>0.333</td>
<td>.003</td>
</tr>
<tr>
<td>Other</td>
<td>7.769</td>
<td>3.993</td>
<td>3.537</td>
<td>0.239</td>
<td>.033</td>
</tr>
</tbody>
</table>

Table 7  Responses by career choice: by year  
(\textit{Number})

<table>
<thead>
<tr>
<th>Year</th>
<th>Family Medicine</th>
<th>Other</th>
<th>P-value*</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>9.765</td>
<td>8.745</td>
<td>.13</td>
</tr>
<tr>
<td>2</td>
<td>9.963</td>
<td>9.600</td>
<td>.76</td>
</tr>
<tr>
<td>4</td>
<td>9.000</td>
<td>7.650</td>
<td>.003</td>
</tr>
</tbody>
</table>

* p-value is a measure of statistical significance. It is the probability that the observations could be due to chance alone (an observation with a p-value of 1 in 20, or p< .05 is considered to be significant and not simply due to chance).