Replacing a structured oral examination by a computer-based assessment for medical master students: from eloquence-based to evidence-based examination?

CONFERENCE PAPER · MARCH 2011
DOI: 10.13140/2.1.5179.6162

5 AUTHORS, INCLUDING:

Bernard Cerutti
University of Geneva
48 PUBLICATIONS 164 CITATIONS

Katherine Blondon
University of Geneva, Genève, Switzerland
22 PUBLICATIONS 134 CITATIONS

Mathieu Nendaz
University of Geneva
121 PUBLICATIONS 978 CITATIONS

Available from: Bernard Cerutti
Retrieved on: 16 March 2016
REPLACING A STRUCTURED ORAL EXAMINATION BY A COMPUTE R-BASED ASSESSMENT FOR MEDICAL MASTER STUDENTS: FROM ELOQUENCE-BASED TO EVIDENCE-BASED EXAMINATION?

B. Cerutti\textsuperscript{1}, A. Galetto\textsuperscript{2}, K. Blondon\textsuperscript{3}, S. Aujesky\textsuperscript{2}, M. R. Nendaz\textsuperscript{2}

\textsuperscript{1}Faculty of Medicine Geneva (SWITZERLAND)
\textsuperscript{2}University Hospital of Geneva (SWITZERLAND)
\textsuperscript{3}Institute for Simulation & Interprofessional Studies Seattle WA (USA)

bernard.cerutti@unige.ch, annick.galetto@hcuge.ch, katherine.blondon@gmail.com, susanne.aujesky@hcuge.ch, mathieu.nendaz@hcuge.ch

Abstract

The last three years (master) of the six-year medical curriculum at the University of Geneva are dedicated to the acquisition of clinical knowledge and competences in clinical settings. In areas such as internal medicine or paediatrics, formal evaluation is usually divided in two phases: an objective structured clinical examination (OSCE) with standardised or real patients, and a structured, oral examination (SOE) based on the resolution of paper clinical scenarios assessed by a principal examiner (a faculty member) and a co-examiner (a chief resident). In both situations, the examiners fill in a grid that allows the computation of a score. Since 2008, the SOE has been progressively replaced by a computer-based assessment (CBA) that evaluates the students’ ability to solve several clinical scenarios. Each scenario is divided into sections of one or more questions. During the test, these sequential sections are given successively, so as to mimic the SOE.

Within the internal medicine setting, we quantified the human resources required for the SOE and for the CBA. We compared the main results of the assessments of the last two years of SOE and the first two years of CBA. We then investigated how these results were correlated with the continuous bedside assessment made during the year in clinical settings.

The reduction in manpower for a single exam easily covered the annual licence fees for the on-line examination program. The internal consistency (Cronbach alpha coefficients) was similar for both types of examination (between 0.8 and 0.9). The number of items was about twice lower for CBA (12.7 vs. 25.9; p<.001) but these items were usually more complex than the dichotomous ones of the SOE grid. More items of the CBA were removed for the final ranking (7.3% vs. 1.0%; p=0.001). Regarding the clinical scenarios, the average score was lower for the CBA (-3.0%; p<.001), and the intra-student variability was similar (p=0.467), but the between-student variability was larger for the SOE (p<.001). The CBA was slightly better correlated with the continuous bedside assessments made in the clinical settings than the SOE (R\textsuperscript{2}=0.181 vs. R\textsuperscript{2}=0.105), especially with the items regarding the association with the clinical documentation and the patient care management.

The shift from oral to computer-based assessment has been broadly accepted by both the students and their examiners. Lack of major changes in reliability and an improved correlation with the continuous assessment in clinical settings support our intent to implement the CBA to other clinical learning units of our curriculum.

Keywords: medicine, computer based assessment.

1 BACKGROUND

The last three years (master) of the six-year medical curriculum at the University of Geneva are dedicated to the acquisition of clinical knowledge and competences in clinical settings. Both teaching and formal evaluation demand a lot of time and personal resources from the university hospital. In areas such as gynaecology, internal medicine, paediatrics, primary care medicine, or psychiatry, formal evaluation is usually divided in several phases that include: an objective structured clinical examination (OSCE) with standardised or real patients and a structured, oral examination (SOE)
based on the resolution of paper clinical scenarios assessed by a principal examiner (a faculty member) an a co-examiner (a chief resident). In both situations the examiners fill in a grid that allows the computation of a score.

In 2007, the internal medicine program directors asked the faculty to evaluate the replacement of the highly time-consuming SOE by a computer-based assessment (CBA). This was aimed at relieving the workload for the clinical staff by reducing both the number of clinical scenarios to be prepared for the formal assessment and the duration of the assessment. After having studied the different products available on the market, the faculty chose the CAMPUS assessment tool, which can be used both for training and formal assessment [1]. This system was chosen in particular because it offered strong guarantees of traceability and security. For example, every examination can be continued if the connection with the server is interrupted. Alternatively, an examinee can continue the examination using another computer if there is a problem with the first one, e.g. a power interruption. The system allows the integration of pictures or videos that can help to improve the content validity of a test [2], and different formats of questions such as: classical multiple choice questions (MCQ) with one correct answer, MCQs with multiple correct answers, fill the gap or provide numerical value questions, point to a region of interest on a picture questions, as well as semi-open, or open questions. The latter however requires a manual grading based on the answer given by each student.

Since 2008, the SOE has been progressively replaced by a CBA that evaluates the students’ ability to solve several clinical scenarios. Each scenario is divided into sections of one or more questions.

2 METHODOLOGY

For the internal medicine exam, the first exam planned for a change of format, we started by investigating the time and personal resources needed for the SOE compared with the CBA, taking into account the exam session only, since the preparation and the analysis were likely to be rather similar.

As a second step, we compared the main characteristics of the SOE and the CBA (reliability with the coefficient of Cronbach, facility of the items, power of discrimination with the biserial correlation coefficient rbis). We used a three-way variance analysis on the scores (SOE or CBA, clinical scenario, and student) to compare the SOE and the CBA, as well as the inter-student and intra-student variability. We also compared the CBA and the SOE results of the students with their continuous bedside assessments made in the clinical settings during their internal medicine clerkship. This continuous assessment includes several items, such as clinical reasoning and patient management, clinical competences, professionalism, clinical documentation, and a general evaluation. We applied a stepwise regression model with the Akaike information criteria to select [3] the relevant items correlated with the SOE or CBA score.

All the analyses were done with TIBCO Spotfire S+® 8.1 for Windows, Palo Alto, CA, USA.

3 RESULTS

3.1 Cost and resources

Compared to an oral exam of 40 minutes per student (two 20-minute oral examinations based on a clinical case; each examination is preceded by 10 minutes of preparation) in a class of 120 students, administering the CBA (60 minutes) saved manpower (9.8 man-days for faculty members, 9.3 man-days for chief residents, 2.5 man-days for external clinical experts, and 4 man-days for office assistants), while the increase in the demand for computer scientists was 2.9 man-days. The reduction in manpower for a single exam easily covered the annual licence fees for the on-line examination program.

During the first year, the time required for the creation of computer based clinical cases, when compared with oral clinical cases, was not formally quantified but was significantly higher than usual (roughly estimated at twice the usual time). Every area of speciality then chose its specific method of developing clinical cases: in some units the MDs took in charge the whole process of development, while in other units the MDs first provided a secretary staff with all the materials (handouts, images, scoring, etc), who helped create the case using the online assessment authoring tool before final verification by the MDs.
3.2 Comparison of SOE and CBA

Table 1 shows the main characteristics of the SOE and CBA exams. The Cronbach alpha coefficients were similar for both examinations. The number of items was about twice lower for the CBA (12.7 vs. 25.9; p<.001), but these items were usually more complex than the dichotomous oral grid items. More items were removed for the final CBA score (7.3% vs. 1.0%; p=.001). Regarding the clinical scenarios, the average score was lower for the CBA (-3.0%; p<.001), the intra-student variability was similar (p=.467), but the between-student variability was larger for the SOE (p<.001).

Table 1: Main characteristics of the last two years of the SOE and first two years of CBA for the internal medicine examination.

<table>
<thead>
<tr>
<th>Examination</th>
<th>Number of students</th>
<th>Number of questions</th>
<th>Cronbach α $\text{100}^{1}$</th>
<th>Precision of the scores</th>
<th>Average rbis $\text{1}$</th>
<th>Average facility %</th>
<th>Number of removed questions</th>
</tr>
</thead>
<tbody>
<tr>
<td>2007 SOE 1</td>
<td>31</td>
<td>25</td>
<td>0.73</td>
<td>7.51</td>
<td>0.37</td>
<td>57.5</td>
<td>1</td>
</tr>
<tr>
<td>2007 SOE 2</td>
<td>31</td>
<td>27</td>
<td>0.77</td>
<td>7.79</td>
<td>0.4</td>
<td>64.4</td>
<td>2</td>
</tr>
<tr>
<td>2007 SOE 3</td>
<td>26</td>
<td>24</td>
<td>0.71</td>
<td>5.84</td>
<td>0.42</td>
<td>83.3</td>
<td>0</td>
</tr>
<tr>
<td>2007 SOE 4</td>
<td>26</td>
<td>30</td>
<td>0.62</td>
<td>5.52</td>
<td>0.21</td>
<td>85.8</td>
<td>0</td>
</tr>
<tr>
<td>2007 SOE 5</td>
<td>24</td>
<td>24</td>
<td>0.71</td>
<td>6.81</td>
<td>0.35</td>
<td>81.9</td>
<td>0</td>
</tr>
<tr>
<td>2007 SOE 6</td>
<td>24</td>
<td>29</td>
<td>0.84</td>
<td>6.78</td>
<td>0.48</td>
<td>70.5</td>
<td>1</td>
</tr>
<tr>
<td>2007 SOE 7</td>
<td>28</td>
<td>32</td>
<td>0.75</td>
<td>5.08</td>
<td>0.35</td>
<td>85.2</td>
<td>0</td>
</tr>
<tr>
<td>2007 SOE 8</td>
<td>28</td>
<td>29</td>
<td>0.83</td>
<td>5.39</td>
<td>0.51</td>
<td>83.9</td>
<td>0</td>
</tr>
<tr>
<td>2008 SOE 1</td>
<td>20</td>
<td>22</td>
<td>0.79</td>
<td>6.04</td>
<td>0.43</td>
<td>86.6</td>
<td>0</td>
</tr>
<tr>
<td>2008 SOE 1</td>
<td>20</td>
<td>30</td>
<td>0.9</td>
<td>5.7</td>
<td>0.78</td>
<td>79.8</td>
<td>0</td>
</tr>
<tr>
<td>2008 SOE 1</td>
<td>21</td>
<td>29</td>
<td>0.81</td>
<td>6.02</td>
<td>0.48</td>
<td>78.8</td>
<td>0</td>
</tr>
<tr>
<td>2008 SOE 1</td>
<td>21</td>
<td>23</td>
<td>0.84</td>
<td>6.82</td>
<td>0.49</td>
<td>79.2</td>
<td>0</td>
</tr>
<tr>
<td>2008 SOE 1</td>
<td>23</td>
<td>26</td>
<td>0.75</td>
<td>6.51</td>
<td>0.34</td>
<td>64.0</td>
<td>0</td>
</tr>
<tr>
<td>2008 SOE 1</td>
<td>23</td>
<td>27</td>
<td>0.84</td>
<td>6.56</td>
<td>0.57</td>
<td>75.1</td>
<td>0</td>
</tr>
<tr>
<td>2008 SOE 1</td>
<td>22</td>
<td>20</td>
<td>0.67</td>
<td>9.08</td>
<td>0.35</td>
<td>57.3</td>
<td>0</td>
</tr>
<tr>
<td>2008 SOE 1</td>
<td>22</td>
<td>18</td>
<td>0.79</td>
<td>7.23</td>
<td>0.45</td>
<td>82.8</td>
<td>0</td>
</tr>
<tr>
<td>2009 CBA 1</td>
<td>111</td>
<td>12</td>
<td>0.86</td>
<td>4.4</td>
<td>0.29</td>
<td>54.4</td>
<td>1</td>
</tr>
<tr>
<td>2009 CBA 2</td>
<td>111</td>
<td>17</td>
<td>0.81</td>
<td>2.49</td>
<td>0.20</td>
<td>79.9</td>
<td>0</td>
</tr>
<tr>
<td>2010 CBA 1</td>
<td>43</td>
<td>15</td>
<td>0.83</td>
<td>2.28</td>
<td>0.27</td>
<td>78.9</td>
<td>1</td>
</tr>
<tr>
<td>2010 CBA 2</td>
<td>43</td>
<td>11</td>
<td>0.92</td>
<td>3.88</td>
<td>0.28</td>
<td>69.2</td>
<td>1</td>
</tr>
<tr>
<td>2010 CBA 2</td>
<td>76</td>
<td>13</td>
<td>0.74</td>
<td>3.13</td>
<td>0.23</td>
<td>75.8</td>
<td>0</td>
</tr>
<tr>
<td>2010 CBA 2</td>
<td>76</td>
<td>8</td>
<td>0.86</td>
<td>5.4</td>
<td>0.22</td>
<td>75.3</td>
<td>3</td>
</tr>
</tbody>
</table>

Mean SOE 24.4 25.9 0.77 6.54 0.44 76.01 0.3
Mean CBA 76.7 12.7 0.84 3.60 0.25 72.25 1.0

$^{1}$ Due to the use of different analysis programs, there may be some slight discrepancies regarding these indicators.
Regarding the CBA, some differences were found between the types of questions: Multiple Choice Questions (MCQ) requesting to select n between m items (average facility 65%; average rbis 0.34) were slightly more difficult (p=0.045) and more discriminatory (p=0.002) than MCQs requesting a single choice (average facility 77%; average rbis 0.18) or semi-open questions (average facility 80%; average rbis 0.22).

3.3 Comparison of SOE and CBA with the bedside continuous assessment

There is a clear correlation between the SOE (p<0.001) or the CBA (p<0.001) and continuous bedside assessments made in the clinical settings, but the error variability is quite important (standard deviation of 8.3 for the CBA and 14.4 for the SOE).

Figure 1: Score at the CBA for every student as a function of the patient care management items (average values) assessed during the continuous training, with grades ranging from 0 to 3 (n=176).

The CBA was slightly better correlated with the continuous bedside assessments made in the clinical settings than the SOE ($R^2=0.181$ vs. $R^2=0.105$), especially with the items regarding the association with the clinical documentation and the patient care.

4 CONCLUSION

As pointed by van der Vleuten and Schuwirth [4], besides classical criteria like reliability and validity, an assessment model should include the acceptability of the method to the stakeholders and the investment required in terms of resources. Other investigators have shown that under given conditions, the results obtained by an oral examination are statistically indistinguishable from the results of a CBA [5], when applied to the same samples of students, on the same topics, and with the same levels of difficulty.

The shift from oral to computer-based assessment has been broadly accepted by both the students and their examiners in our institution. The reduction in manpower for a single exam, compared to SOE, easily covered the annual licensing fees of CBA. We emphasize the importance of constructing CBA test items of high quality, since unlike SOE, this format does not allow the examiners to clarify ambiguous questions. As the type of questions were not completely the same in diverse studies, direct comparisons are difficult. However, in our study, the lack of major changes in reliability and a slightly
improved correlation with the continuous assessment in clinical settings support our intent to implement CBA to other clinical learning units of our curriculum.

REFERENCES


