

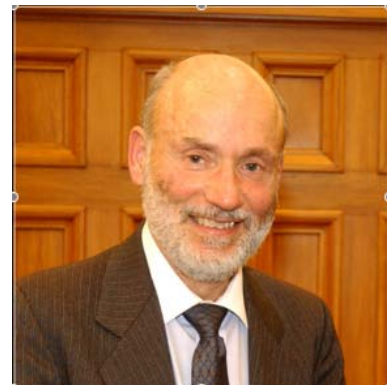


A Most Unusual Examination: Progress Testing at the University of Otago Medical School

Author

Peter Schwartz, University of Otago

The Otago Medical School moved from reporting assessment results as grades to reporting in the categories of 'distinction', 'pass' or 'fail'. In response to this change students expressed concern about how they would know precisely how they were performing. The introduction of progress testing has provided some of this requested feedback to learners since the change in the method of reporting assessment results.



Introduction and Background

'But how will we know how we're doing?!' This was one of the most common initial reactions from students when the University of Otago Medical School announced that, from 2002, all reporting of student performance would be solely as 'distinction', 'pass' or 'fail' instead of the marks and grades that had been used until then. The change was made partly to eliminate the spurious aura of accuracy and precision implied when assessment results are reported as marks or grades^{1,2} and, just as for other medical schools that made this change,^{3,4} partly to discourage academic competition and encourage cooperation among medical students. Still, the students complained about the change. Faculty assured the students that measures would be put in place to provide the feedback that the students so desperately wanted (and that has been shown to be so important to learning^{5,6}).

It was during this interval of disquiet that I remembered something that I had seen and read about several years earlier: the progress test. This was a novel type of examination that was developed independently by two medical schools (one in the United States,⁷ the other in the Netherlands⁸) during the 1970s. The concept has since been adopted by a number of other medical schools, most notably McMaster in Canada, the originator of the full problem-based learning medical curriculum.⁹

The most unusual feature of the progress test is that an identical test is sat by ALL students in the medical school at about the same time. The students sit a different but equivalent test several times each year (usually two to four tests per year). Questions are usually either multiple choice (University of Missouri-Kansas City,⁷ McMaster⁹) or true-false (University of Limburg⁸), but some medical schools are experimenting with short answer questions.¹⁰ Each test is prepared from a blueprint that specifies a

certain percentage of questions from every main discipline or domain, both basic medical sciences and clinical disciplines. After a test is completed, each student receives complete, personalised feedback on performance, overall and in the various disciplines and subdisciplines, along with comparative figures for the student's own class and all the other classes in the medical school. The expectation is that, on any test, performance overall and in the individual disciplines will be better on average for classes that are further along in the course, and, for any class or individual, performance will be better as the class or student moves through the course. For whole classes, these expectations are largely met, especially for the scores in the clinical components. There can be somewhat more variability for individual students and some medical schools recommend remedial measures^{9,11} (or may even delay progression) for students whose performance on these tests is consistently substantially below that of their peers. Others, however, use the tests only as formative measures for the students' own benefit.

For the students, one major advantage of the progress test is that it is so wide ranging in content that there is no sense in (and no expectation of) preparing specifically for the test. It can be scary for new students faced with clinical questions comprising vocabulary that they don't even recognise, but they soon get used to this and they can at least see the sorts of things that they will be expected to be familiar with later in the medical course. For the faculty, an advantage is that the test does not interfere with either the style or the content of teaching in individual components of the curriculum. The objectives tested reflect the overall final product of the medical course rather than discipline, module, or block objectives. (It was this advantage that prompted McMaster to adopt progress testing despite the apparent incongruity of such a test being used in a problem-based learning curriculum.⁹)

Introducing Progress Testing at Otago Medical School

I had seen (and not been impressed by) an example of a progress test comprising true-false questions. As the Faculty was considering how to respond to students' requests for feedback in the new 'distinction/pass/fail' reporting system, however, I obtained some examples of the multiple-choice question (MCQ) tests that were administered by the University of Missouri-Kansas City (UMKC). To me these looked much better than the true-false tests. The UMKC administered tests four times per year. Each test consisted of 400 questions, and each test had 30% of the questions in the discipline and subdisciplines of medicine (including public health, preventive medicine and psychiatry), 15% of the questions in each of paediatrics, obstetrics & gynaecology, and surgery, and 25% in the basic sciences (5% each in anatomy/histology, biochemistry, microbiology, pharmacology, and physiology – pathology was incorporated into the clinical disciplines). The questions for each test were selected from a bank of some 16,000 MCQs (!) prepared by faculty members. Most of them tested only factual material, but a good proportion required at least some interpretation or application.

Although the questions had not been prepared locally (and some were distinctly USA-specific), I felt that they represented a good selection of items from the 'universe' of facts and concepts that could be relevant to newly graduated doctors. They appeared to represent a good basis for developing a scheme in which students at Otago would be able to obtain feedback on at least their developing funds of knowledge as they moved through the medical curriculum, and to identify areas of strength and weakness. I recognised that this was only one component of what the students were expected to achieve in the medical course, but it was an important one.

Through the kind cooperation of one of the chief architects of progress testing at UMKC, I was able to gain access to a number of the recent tests that had been sat by the students of UMKC. During 2003, I

proposed to the Faculty assessment committee that we run a voluntary pilot study with the fifth year students only, administering a single progress test several weeks before the end-of-year examinations. For this purpose, we would use exactly the same test as one that had recently been administered at UMKC. The committee agreed with this proposal and the pilot study was subsequently run. The test was run in two parts, 200 questions in each part, up to three hours allowed to complete each part. The students were accommodated in a lecture theatre, the questions were on paper, and answers were entered onto optical mark readable sheets. After reading at Otago's computing centre, the answers were scored and analysed by my contact at UMKC, who also provided files from which we could print out individual score reports for distribution to the students. In the event, 77% of the fifth year class sat the test. The trial ran well and, on a questionnaire afterward, 82% of the students who sat the test were in favour of similar tests being offered in the future, despite some misgivings about the format and content of some of the questions. This result was encouraging enough that I suggested to the assessment committee that we run another, slightly bigger trial in 2004.

Full Progress Testing at the University of Otago

The assessment committee disagreed. Instead, they asked that from 2004 I start a full run of progress testing for all students in years 2 through 5 at the medical school, two tests per year, the first about halfway through first semester, the second about halfway through second semester. As I had little or no technical or administrative support, this posed something of a challenge. Again I was fortunate to have the full cooperation of my contact at UMKC, who agreed to supply files of their recent tests comprising about a quarter of their bank of items and to analyse all of our results and supply files with the individual feedback reports. (All of this, by the way, was done free of charge and with no strings attached.) By obtaining designated times in the timetables for the various classes, I was able to test over 900 students twice each year by the method used in the pilot study (in lecture theatres, pencil and paper, two three-hour sittings per test, each sitting for 200 of the 400 test items).

For both of the tests in 2004 and the first test in 2005, I used tests that were identical to those used recently at UMKC so that we could assess comparability between our students and those at UMKC. There was consistently only a slight difference (usually within about 5%, with mean scores in the clinical disciplines sometimes higher at UMKC and sometimes higher at Otago, while scores in the basic medical sciences were consistently higher at UMKC among the advanced clinical students, probably because of their repeated exposure to questions from the basic sciences in their previous progress tests). From the second test in 2005, I used questions from the UMKC bank for our tests after discarding the most obviously USA-specific ones (Otago students invariably commented if a question about Rocky Mountain Spotted Fever appeared in a test or if an answer depended on an awareness of patterns of diseases in the United States).

From 2004 through 2006, progress testing at Otago ran in this fashion. The tests were 'compulsory', but no penalties were imposed on students who failed to sit and no one other than the student him- or herself saw the individual results. Nearly 100% of students sat the tests. Results were reported as 'raw' percentages correct. There was no correction for 'guessing'. The pattern of results was as expected: higher scores on a given test by classes further into the course and increasing scores by any class as it sat further tests. (See Figures [1](#) and [2](#) for recent examples.) It was left to individual students to respond to their own results, although they were encouraged to seek advice or help if they found that their results were consistently well below those of their classmates, either overall or for particular disciplines or subdisciplines. As part of the individual feedback, each student received a graph showing his or her total

scores on all tests sat to date, superimposed on a zone showing the mean \pm 1 standard deviation for his or her class for the same tests. (See [Figure 2](#) for an example.)

Figure 1. Mean scores by Otago Medical School classes in years 2 through 5 on first progress test for 2009, total and for medicine, paediatrics, obstetrics & gynaecology, surgery, and combined basic sciences.

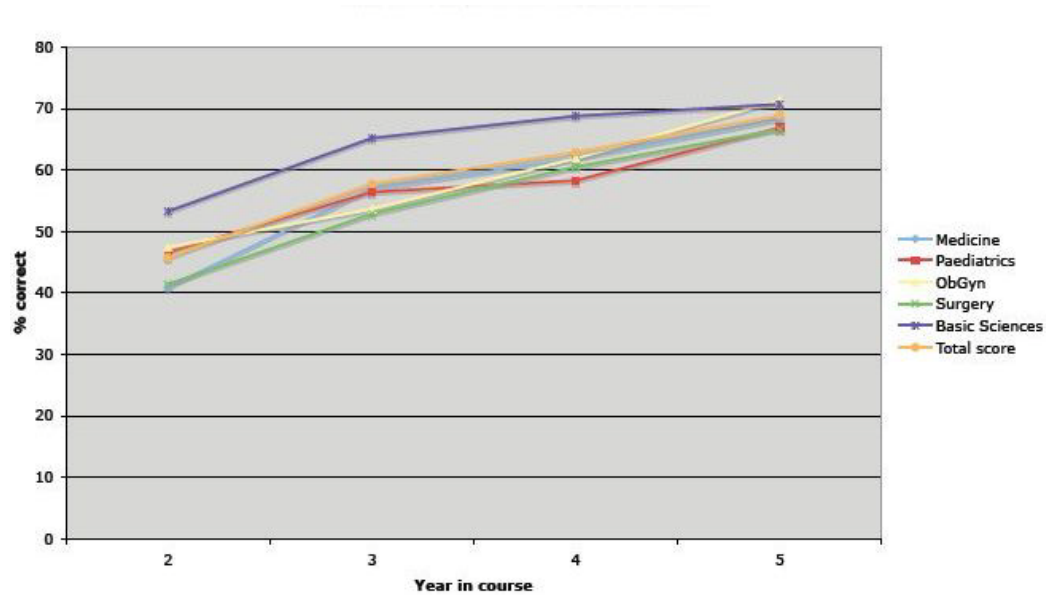
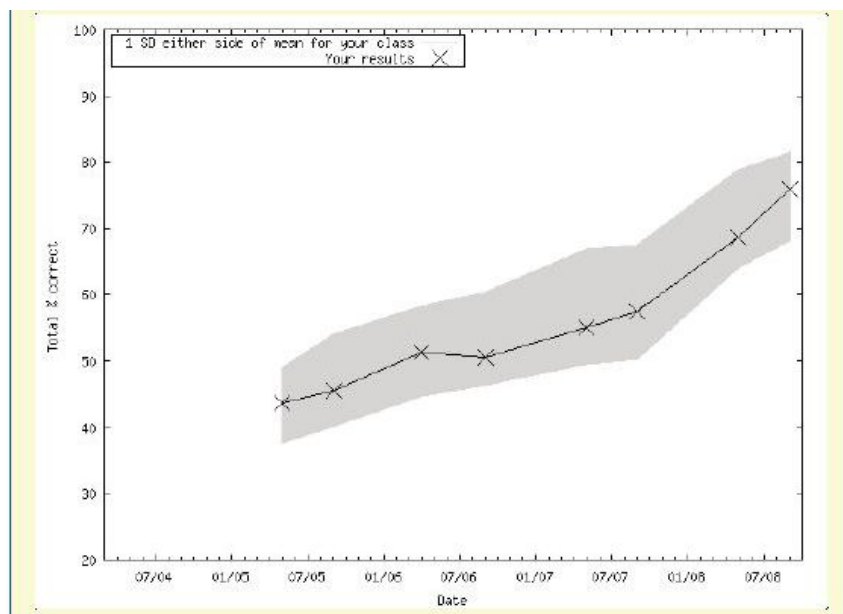


Figure 2. Example longitudinal total scores for progress tests for fifth year class 2008 plus superimposed results for one student. The grey zone shows the mean total scores \pm 1 SD for all eight progress tests sat by the fifth year class of 2008 (two tests each year 2005-2008). The means for the class are at the centre of the grey zone. Superimposed (as Xs joined by a line) are the total scores for one student in the class on those tests. His/her total scores were consistently near the mean for the class.



To me the most striking and unexpected observation from these first years of progress testing at Otago was the very high correlation between students' total scores on the last progress test in fifth year and their scores in the total written component of the fifth year final examination. This was despite the facts that:

- the progress test MCQs bore no resemblance to the fifth year exam questions (which were extended matching multiple choice and short answer types)
- the content of the progress test questions was not selected to match the objectives of the fifth year exam
- students did no preparation whatever for the progress test, while they most certainly did for the fifth year exam.

From the pilot study in 2003 through 2006, the correlation coefficient between results on the last progress test in fifth year and performance in the fifth year written final examination was between 0.74 and 0.79. I reached the stage of being able to tell students that, if their progress test results kept pace with the rest of their class and if they were able to score over about 65% overall on the last progress test before the fifth year exams, they could be confident that they would almost certainly pass the fifth year written exam.

Paradoxically, this ongoing consistently high correlation was a bit disturbing. I had hoped that it might fall somewhat as students discovered from their progress tests the areas in which they had deficiencies and proceeded to rectify these before the final exams. This did not seem to be happening.

Evaluating and Modifying the Progress Test

During 2006, the medical education support team at Otago undertook an evaluation of the progress tests. Students reported being generally satisfied with the tests (two-thirds of respondents favoured retaining the tests), but they were concerned by the amount of time required to complete the tests. Clinical teaching staff and some students objected to the encroachment of the tests on time in clinical attachments. In response to these concerns, since 2007 the progress tests have been administered as computerised tests that students can sit at any time over a two week interval from any computer with internet access. The number of questions per test has been halved, while still retaining the same blueprint. (In a split test experiment, I showed that cutting the numbers of questions had little effect on the predictive ability of the test at fifth year level.)

The change in format of the test has removed the previous complaints from teachers and students. In a more recent evaluation by the medical education support team, however, there are now some new ones. About 15% of students admit to looking up answers (on the internet or in books) before answering. While this might arguably be considered a form of learning, the students have repeatedly been reminded that the main purpose of the test is to gauge the student's developing fund of knowledge in relation to that of the other students in his or her class. Some students are complaining about this apparent 'cheating' by others. And some students now say they don't take the test so seriously because it has so much flexibility and because they're not sitting it in an exam type setting. I guess I can't win no matter what I do. In any event, the patterns of performance by different classes on individual tests and by the same class over time are similar to what they were before. But the correlation between performance on the last progress test in fifth year and the written final exam score has fallen considerably since the change in 2007 (to about 0.65 in 2007 and further to about 0.53 in 2008). It would be nice to think that this was because students were using information from the progress tests

constructively, but the pattern of results suggests it is more likely to be the effect of some students looking up answers and others not taking the progress test so seriously (see Figures 3 and 4 for scatter diagrams from 2005 and 2008).

Figure 3. Scatter diagram showing total score on last progress test in fifth year vs total written exam score in fifth year final exam, 2005. Correlation coefficient (r) = 0.76.

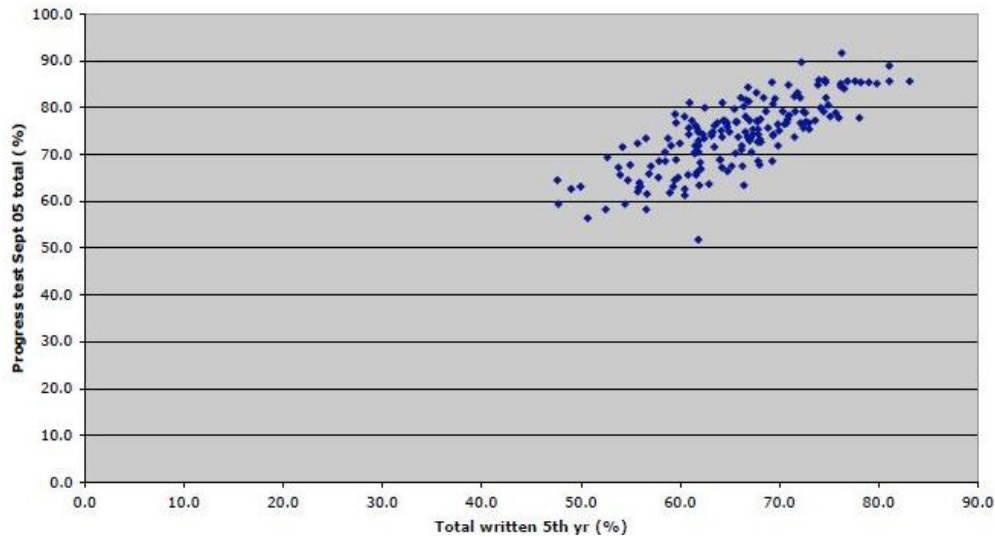
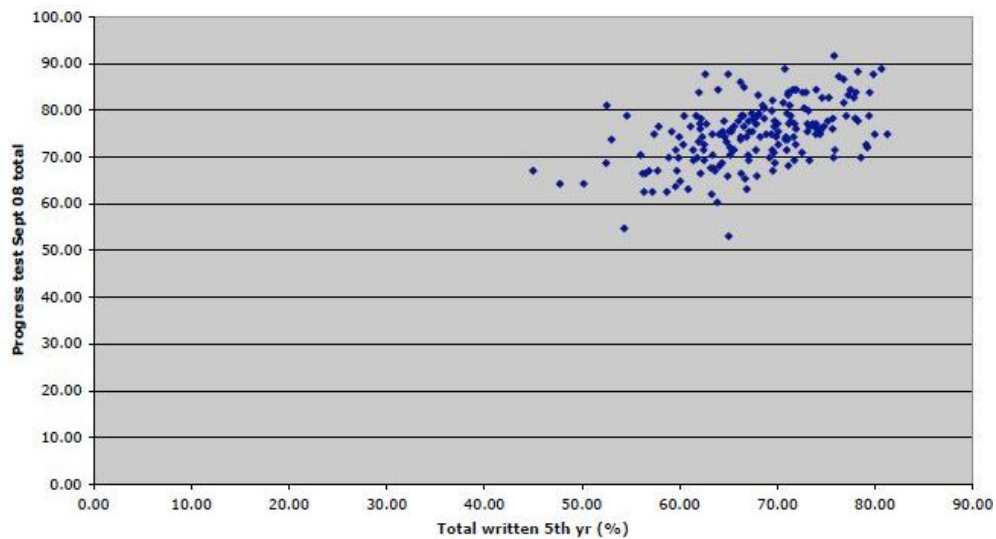


Figure 4. Scatter diagram showing total score on last progress test in fifth year vs total written exam score in fifth year final exam, 2008. Correlation coefficient (r) = 0.53.



For the present, progress testing continues as it has since the start of 2007. But anything might happen. Watch this space!

Note added at proof stage

As the fifth year written exams for 2009 approached, I sent individual e-mail messages to all fifth year students pointing out what had happened in 2007 and 2008 and reminding them that the most valid feedback from the progress test would be obtained if they sat the test as intended: taken seriously, without specific preparation, and without looking up answers before entering their own answers during the test. I asked them to sit the upcoming test in this fashion. Whether this request was responsible or not, for 2009 the correlation coefficient between overall scores in the last progress test and results from the fifth year written final exam rose to about 0.70, nearly as high as the coefficients had been between 2003 and 2006. I'll be interested to see what happens in 2010!

References

1. Kohn, A. (1999) *The Schools Our Children Deserve*, Boston: Houghton Mifflin.
2. Skurnik, L.S. and Nuttal, D.L. (1968). Describing the reliability of examinations, *The Statistician*, 18:119-28.
3. Hurt, A. (2005, Sept) Is medical school about passing or winning? *The New Physician*, 54(6), <http://amsaweb701.rd.net/AMSA/Homepage/Publications/TheNewPhysician/2005/tnp75.aspx>
4. (accessed 27 July 2009).
5. Niedowski, E. (2002) Marking a new era, Hopkins drops grades, *The Baltimore Sun*, 11 October 2002, <http://www.papillonsartpalace.com/johns.htm> (accessed 27 July 2009)
6. Simpson, M.A. (1972) *Medical Education. A Critical Approach*. London: Butterworths.
7. Wood, D.F. (2007) *Formative Assessment*, one of 29 booklets in the ASME series Understanding Medical Education. Edinburgh: Association for the Study of Medical Education.
8. Arnold, L. and Willoughby, T.L. (1990) The quarterly profile examination, *Academic Medicine*, 65:515-6.
9. Van der Vleuten, C.P.M., Verwijnen, G.M. and Wijnen, W.H.F.W. (1996) Fifteen years of experience with progress testing in a problem-based learning curriculum, *Medical Teacher*, 18:103-9.
10. Blake, J.M., Norman, G.R., Keane, D.R., Mueller, C.B., Cunnington, J. and Didyk, N. (1996) Introducing progress testing in McMaster University's problem-based medical curriculum: psychometric properties and effect on learning, *Academic Medicine*, 71:1002-7.
11. Rademakers, J., ten Cate, Th.J. and Bär, P.R. (2005) Progress testing with short answer questions, *Medical Teacher*, 27:578-82.
12. Blake, J.M., Norman, G.R. and Smith, E.K.M. (1995) Report card from McMaster: student evaluation at a problem-based medical school, *Lancet*, 345:899-902.



This work is published under the [Creative Commons 3.0 New Zealand Attribution Non-commercial Share Alike Licence \(BY-NC-SA\)](https://creativecommons.org/licenses/by-nc-sa/3.0/). Under this licence you are free to copy, distribute, display and perform the work as well as to remix, tweak, and build upon this work noncommercially, as long as you credit the author/s and license your new creations under the identical terms.