

ORIGINAL RESEARCH

PROGRESS TESTING: EVALUATION OF FOUR YEARS OF APPLICATION IN THE SCHOOL OF MEDICINE, UNIVERSITY OF SÃO PAULO

Eliane R. Tomic, Milton A. Martins, Paulo A. Lotufo, and Isabela M. Benseñor

Tomic ER, Martins MA, Lotufo PA, Benseñor IM. Progress testing: evaluation of four years of application in the School of Medicine, University of São Paulo. Clinics. 2005;60(5):389-96.

BACKGROUND: Progress testing is a longitudinal tool for evaluating knowledge gains during the medical school years.

OBJECTIVES: (1) To implement progress testing as a form of routine evaluation; (2) to verify whether cognitive gain is a continuous variable or not; and (3) to evaluate whether there is loss of knowledge relating to basic sciences in the final years of medical school.

METHODS: A progress test was applied twice a year to all students from 2001 to 2004. The mean percentage score was calculated for each school year, employing ANOVA with post hoc Bonferroni test evaluation for each test.

RESULTS: Progress testing was implemented as a routine procedure over these 4 years. The results suggest a cognitive gain from first to sixth year in all eight tests, as a continuum (P for trend $< .0001$). Gain was found to be continuous for basic sciences (taught during the first 2 years), clinical sciences ($P < .0001$), and clerkship rotation ($P < .0001$). There was no difference between the performance of men and women.

CONCLUSION: Progress testing was implemented as a routine, applied twice a year. Data suggest that cognitive gain during medical training appears to be a continuum, even for basic science issues.

KEYWORDS: Progress testing Longitudinal test. Assessment. Teaching. Medical graduation.

One of the most important aspects of medical competence and clinical reasoning is the need for physicians to acquire a great capacity to accumulate information with an organized method.¹⁻³ This ability must be taught from the outset at the medical school. In addition, if so much information is needed for a student to become a good professional, medical schools ought to create specific tools to evaluate the acquisition of knowledge during the school years.

The quantification of knowledge gain over the course of the medical school years is a challenge. Recently, the longitudinal tool of progress testing has been shown to be suitable for application to curricula involving problem-

based teaching in a number of medical schools. Progress testing was especially developed to measure cognitive skills in a problem-based learning universe. However, more recent data suggest there is no dilemma regarding its use in a school with a nonproblem-based learning curriculum.⁴ In fact, a similar testing procedure, named the *Quarterly Profile Examination*, has been developed in parallel to progress testing at the School of Medicine at the University of Missouri, in Kansas City.⁵

In 2001, the School of Medicine at the University of São Paulo decided to apply progress testing twice a year to all students from first to sixth year as a way of evaluating the learning of cognitive skills in a nonproblem-based curriculum. The present study attempts to describe the implementation of progress testing at our school. Adaptations we have made over the last 3 years may be useful as a model for future applications of progress testing in other schools with similar curricula. In this study, we have also

Department of Internal Medicine, Hospital das Clínicas, Faculty of Medicine, University of São Paulo – São Paulo/SP, Brazil.

Email: isabensenor@hcnet.usp.br

Received for publication on May 16, 2005.

Accepted for publication on July 19, 2005.

tried to address 3 specific issues: (1) to implement progress testing as a form of routine evaluation; (2) to verify whether cognitive gain is a continuous variable or not; and (3) to evaluate whether there is loss of knowledge relating to basic sciences in the final years of medical school.

METHODS

The curriculum of the School of Medicine at the University of São Paulo is divided into 3 cycles, each one lasting 2 years. The first 2 years include basic sciences (anatomy, physiology, cell biology, and others); the next 2 years relate to clinical sciences; the final 2 years include clerkship rotation, basically in outpatient clinical facilities, general wards, and emergency rooms. Students rotate through internal medicine, pediatrics, surgery, obstetrics, and gynecology.

Characteristics of progress testing at School of Medicine, University of São Paulo

Progress testing was introduced to our students in the first semester of 2001. We decided to apply the test only twice a year until we were more familiar with it. In every semester, the students have a class-free day dedicated to evaluating the medical course with staff members. This evaluation is in seminar-format, and held during the morning, while the progress test is given during the afternoon.

In its first application, the test consisted of 130 questions on basic sciences, clinical sciences, and clerkship rotation. The test format was then restructured to include 100 questions subdivided into 33 questions about basic sciences, 33 about clinical sciences, and 34 about clerkship rotation issues. The number of questions relating to each discipline was calculated on the basis of the number of hours allotted to that discipline in the school curriculum. As the test is applied 1 month before the end of the school semester, we expected, for instance, that a student in the fourth semester would be able to answer all the questions from the first, second, and third semesters, and 80% of the questions relating to the fourth semester.

In Brazil, the most common type of question used in tests to evaluate students is a multiple-choice question with 5 alternatives. Because most staff members are familiar with devising this questions in this format, and students are trained to answer such questions, we decided to use only this type. We did not include an alternative in the "I do not know" format, because there is no tradition for the use of this type of question in Brazil. Students would be afraid to say that they did not know how to answer questions about disciplines they had already completed. We did not

penalize students who answered questions incorrectly. So, the scores were calculated by adding the number of correct answers in the test, and the results were subsequently presented as percentages. We tried to adapt the format of the test to match the format of our routine tests as a way of decreasing possible dissatisfaction with the implementation of a new test.

The questions used in the test were devised by staff members in all the departments of the medical school and were selected and compiled by the team responsible for organizing the progress testing. The questions could include figures or graphs and had to be restricted to issues that were fundamental to the respective discipline for a future physician.

The first 4 applications of the progress test were nonmandatory. Initially, we tried to discuss its importance with the students. The discussion focused on the fact that the test was not being used for promotional objectives, since each discipline has its particular form of evaluating students. After 2 years of testing, we decided the test should become compulsory and absent students would have to justify their absence.

After results are released, students have 7 days to register complaints by e-mail. All complaints are analyzed, and the questions with which the students had greatest problems are disregarded.

Degree of difficulty and discrimination in the tests

In the first 4 tests, we did not include all disciplines either when they failed to comply with the deadline or because the quality of questions was unacceptable. However, from the second test onwards, at least 90% of all disciplines were included. There was no difference in the degree of difficulty of the questions used in the tests over the years (mean degree of difficulty). However, the questions gradually became more discriminative. Thus, in the last 4 tests, the questions had good discrimination power in clinical sciences and clerkship rotation, while still needing some improvement in basic sciences.

We continue to request new questions every semester with the objective of creating a good-quality question bank. After 3 years, we have at least 2000 good quality questions from all disciplines. However, we are still collecting new questions during the preparatory period before test application.

Statistics

Mean and standard deviation scores for all students were calculated according to gender and school year for each

occasion on which the test was applied. Mean scores for each discipline were calculated for all the students according to their school year. Mean scores for the basic course, clinical course, and clerkship rotation years were also calculated. Comparisons between the mean scores for students from first to sixth years on each test occasion were made using the ANOVA test with post hoc Bonferroni test evaluation.

RESULTS

Table 1 shows the student attendance rates in the 8 tests, according to school year. In the first 4 applications, attendance was not compulsory. After the test became manda-

tory, attendance of first-year students increased, but attendance of final-year students was lower than in the previous tests. Except for the first test, in which attendance was higher for women than for men ($P < .0001$), in the other 3 tests for which attendance was not compulsory, there was no difference between genders.

Figure 1 shows mean percentage scores according to undergraduate year and test number. Sixth-year students had a worse performance than did fifth-year students in 2 of the 4 tests with compulsory attendance. However the gain in knowledge was still significant (P for trend $< .0001$) in all tests.

Figures 2, 3, and 4 show the mean percentage scores relating specifically to basic sciences, clinical sciences, and clerkship rotation issues for students from first to sixth year

Table 1 - Student attendance in the 8 applications of progress testing, according to school year

School year	Test 1 n (%)	Test 2 n (%)	Test 3 n (%)	Test 4 n (%)	Test 5 n (%)	Test 6 n (%)	Test 7 n (%)	Test 8 n (%)
First	30 (17.1)	12 (6.7)	17 (9.4)	10 (5.6)	169 (94.9)	167 (94.4)	164(93.7)	147(83.1)
Second	74 (40.9)	26 (14.4)	16 (8.7)	22 (11.9)	108 (56.5)	132 (71.0)	181(98.9)	157(86.2)
Third	62 (34.3)	61 (33.7)	59 (32.2)	43 (23.9)	147 (83.1)	116 (64.8)	136(75.1)	117(64.3)
Fourth	62 (36.5)	55 (32.4)	86 (44.5)	40 (21.7)	172 (90.1)	136 (76.4)	141(76.2)	85(45.7)
Fifth	73 (40.1)	53 (29.1)	64 (37.9)	39 (23.3)	172 (83.7)	131 (71.6)	132(72.9)	108(59.0)
Sixth	113 (64.6)	28 (16.0)	83 (46.1)	49 (28.7)	182 (96.2)	106 (61.9)	146(77.2)	122(63.9)
All	414 (38.7)	235 (22.0)	325 (29.9)	203 (19.1)	898 (83.8)	747 (69.2)	759(69.4)	736(66.5)

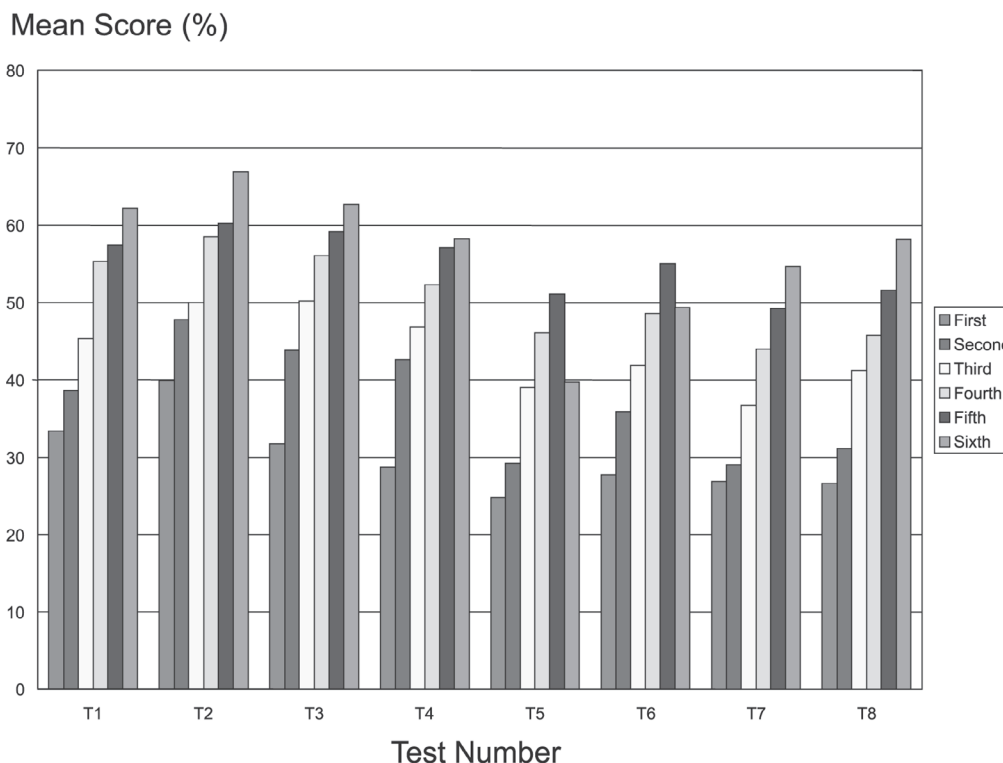


Figure 1 - Mean scores for all questions, for students from first to sixth year* according to occasion on which the test was applied (tests 1-8) (for all tests, P for trend $< .0001$). *For each sequence of columns, the first represents first-year students; the second, second-year students; the third, third-year students; the fourth, fourth-year students; the fifth, fifth-year students; and the sixth, sixth-year students.

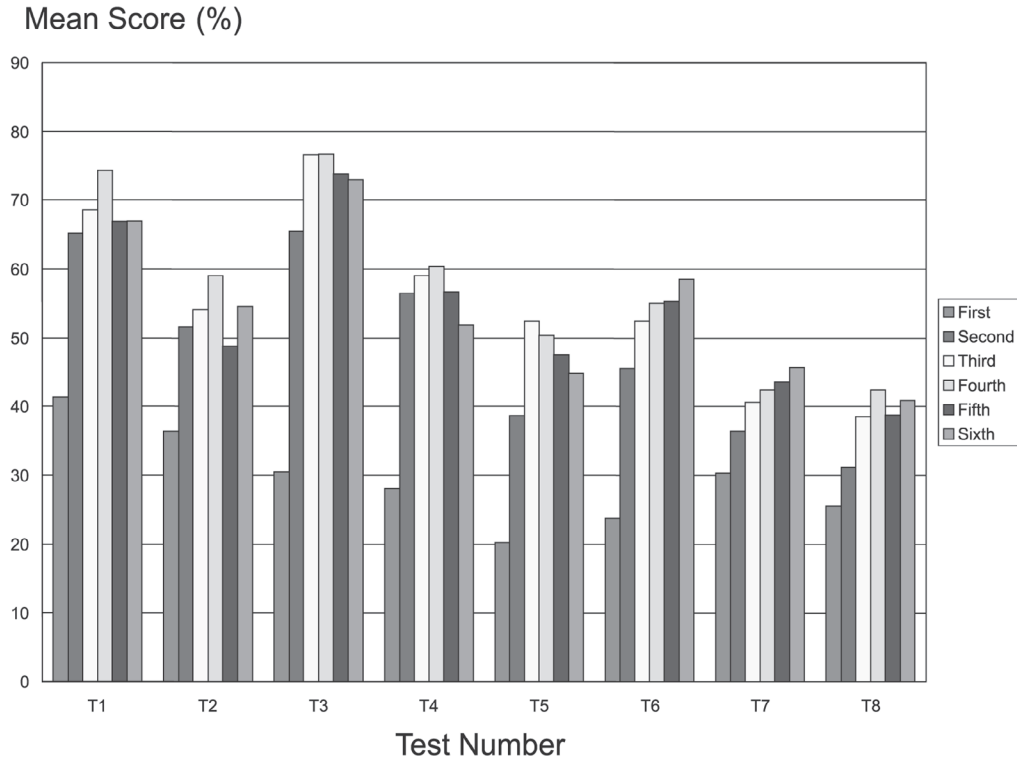


Figure 2 - Mean scores (%) for basic science questions, for students from first to sixth year, according to occasion on which the test was applied (tests 1-8). (tests 1, 3, 5, 6, 7 and 8 $P < .0001$; test 2, $P = .04$; test 4, $P = .03$). *For each sequence of columns, the first represents first-year students; the second, second-year students; the third, third-year students; the fourth, fourth-year students; the fifth, fifth-year students; and the sixth, sixth-year students.

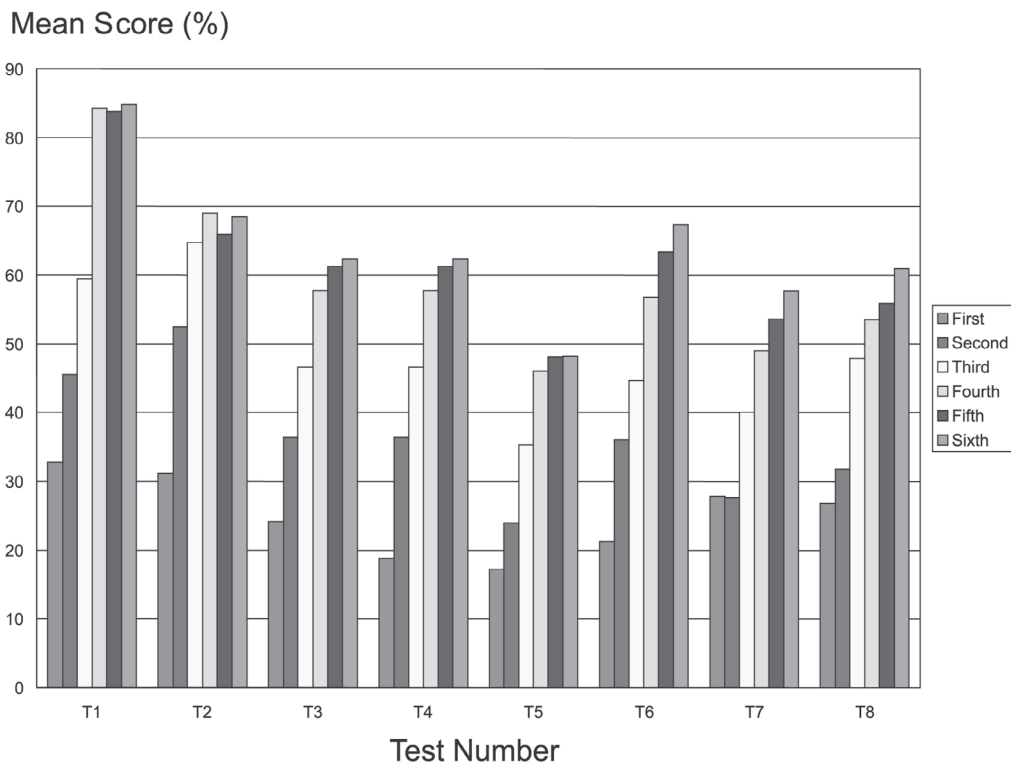


Figure 3 - Mean scores (%) for clinical science questions, for students from first to sixth year, according to occasion on which the test was applied (tests 1-8). (test 1 to 8, $P < .0001$). *For each sequence of columns, the first represents first-year students; the second, second-year students; the third, third-year students; the fourth, fourth-year students; the fifth, fifth-year students; and the sixth, sixth-year students.

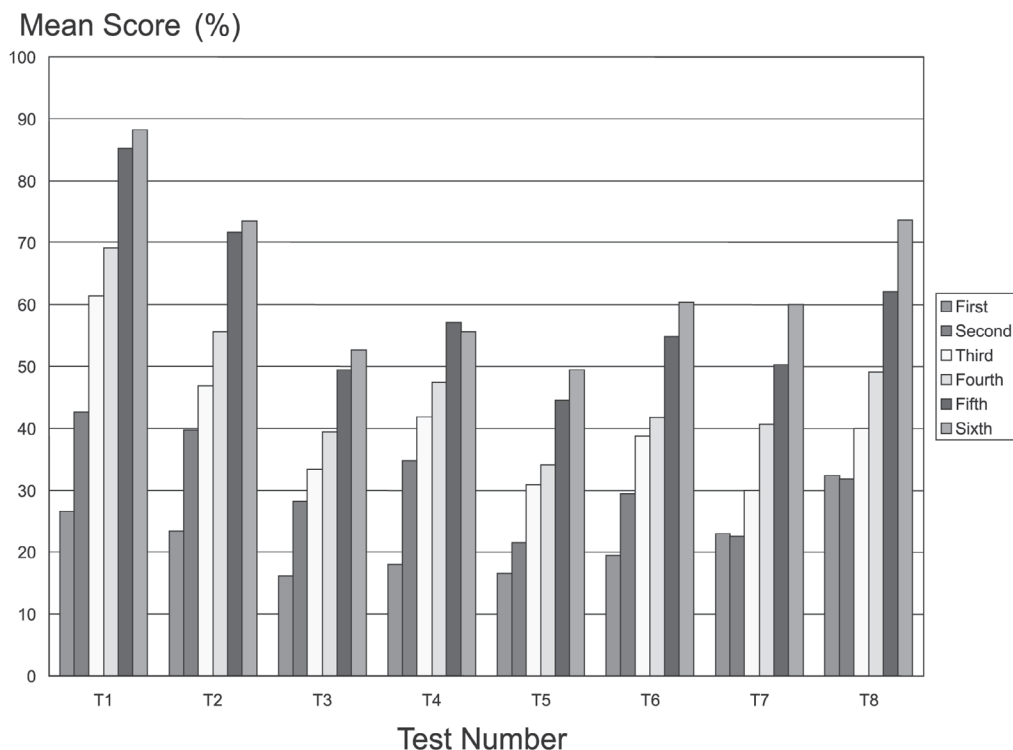


Figure 4 - Mean scores (%) for clerkship rotation questions, for students from first to sixth year, according to occasion on which the test was applied (tests 1-8). (For all tests, $P < .0001$). *For each sequence of columns, the first represents first-year students; the second, second-year students; the third, third-year students; the fourth, fourth-year students; the fifth, fifth-year students; and the sixth, sixth-year students.

(P for trend $< .001$). Except for basic sciences in the second test ($P = .04$) and fourth test ($P = .03$), the results suggest progressive cognitive improvement over the course of the undergraduate years.

There was no difference in mean percentage score between men and women (data not shown).

DISCUSSION

Our results suggest a progressive cognitive gain from first to sixth year in all tests. Even for basic sciences, the data suggest a possible continuous cognitive gain over the entire course of undergraduate years. Men and women had the similar performance. Progress testing seems to be a good longitudinal tool for evaluating gain of knowledge at the School of Medicine at the University of São Paulo. After eight applications, the test has been incorporated into the routine of each semester.

The evaluation of progress testing at the School of Medicine in the University of São Paulo differed in a few aspects from the evaluations at McMaster University (Canada) and the University of Maastricht (Netherlands).^{6,7} The first difference is that we did not use “I don’t know” answers, because there is no tradition in Brazil of using this type of alternative.⁸ Our students would probably be

afraid to answer “I don’t know,” thinking that they could be penalized for doing so. Consequently, the scores from the progress tests at our school were calculated using only the questions answered correctly.

However, the major difference is that in our school, each discipline was responsible for devising questions for progress testing, and the number of questions was calculated on the basis of the number of hours allotted to each discipline in the school curriculum. This is very different from the University of Maastricht, where the selection of questions was on the basis of the blueprint. Therefore, even though the mean scores of between 50% and 60% obtained at our school are very similar to the results from other schools like Maastricht (mean score of 58%, calculated from the correct answers alone), it is possible that the results are not comparable.^{8,9}

The other difference is that since we do not have any experience with “true or false” questions, we used multiple-choice questions, as used by McMaster University. Through this, we aimed to conduct the progress testing in the manner used for all other evaluations at our school. Results also show that overall knowledge increased uniformly with time, as the training progressed from first to sixth year; this result is similar to that observed at the University of Maastricht and other institutions.^{10,11}

We had no difficulty in adapting progress testing to a medical school with a traditional curriculum. Although progress testing was created for evaluating schools using problem-based learning, further comparison of student performance between medical schools with and without problem-based learning has only shown small differences. There was no difference in cognitive performance between schools using either type of curriculum. When the comparison was divided into 3 categories (basic, clinical, and social sciences), few differences were found.^{12,13} Students not using problem-based learning scored better in basic sciences, while students that used problem-based learning scored better on the social sciences.^{12,13}

It is possible to speculate about our data. The change from optional to compulsory brought some modifications in student attendance according to undergraduate year. When the test was optional, students in the final years, who were more concerned about using the test as training ground for the medical residence admission exams, presented the highest attendance. Some first-year students feared that the test results could interfere with their progression to the next school year, and consequently did not come to the first tests. After the test became compulsory, several sixth-year students came to the test merely to register their attendance, and either did not answer any questions in the tests, or answered only the questions relating to clinical rotation issues. This could explain the lower score among sixth-year students in the fifth and sixth tests. However, the gain of knowledge appears to be significant. Comparing the curves from the first 4 tests, for which attendance was not compulsory, with those for the last four tests, for which attendance was mandatory, there is no great difference. The lower scores for the last four tests can be attributed to the change in student attendance, but can also be explained by the greater discriminative power of the questions, which improved over the 4 years of testing. Except for the first test, attendance by men and women was the same.

We understand that having no apparent loss of cognitive gain relating to basic sciences over the school years is good news. One possible explanation for these results is that some basic science issues are revisited during the clinical course and hospital rotations.

In Brazil, more women than men apply for the selection exams to enter medical school each year. However, at medical school we have more men than women. As men

and women have similar performances regarding cognitive knowledge in the medical school, this probably indicates that we may have some bias in the entrance exams.

Implementation of progress testing at our school was more difficult than expected. Some complaints by the students related to their fear that we were creating a new form of evaluation that could be used in assessments for progression to the next school year. Some students also said that the test stimulated competition between them. Most of the students only accepted the importance of the test when the results were shown to them for the first time. Now, after 4 years, the test has become part of the routine at our school. Students demand detailed comments on all the alternatives for each question after each test, as a way of using the test to help them in their learning.

Elsewhere in Brazil, progress testing has only been applied regularly at the Federal University of São Paulo, for 4 consecutive years once per year. The results also suggest progressive gain of knowledge over the years, even for questions about basic sciences.¹⁴ Now, several schools are implementing progress testing, for example at the Federal University of Minas Gerais, the School of Medicine of Marília, and the State University of Londrina. This will enable future exchange of information.

There are a number of limitations associated to the way in which we implemented progress testing, and these need to be considered. The format of the test, with only 100 questions, means that some disciplines are represented only by a single question, or are competing for their question with several other disciplines with small representation in the medical curriculum. If the question is unrepresentative or inadequate, the evaluation of the discipline could produce biased results. Another point to be discussed is that the data might not be representative of all the students because of the lower attendance in the first four tests when the test was not mandatory. Even in the last 4 applications of the test, the attendance ranged from 66% to 84%, which allows for the possibility of some kind of selection bias.

The data suggest that cognitive gain appears to be a continuum over the course of undergraduate years, even for basic sciences. Therefore, progress testing could be used as a further instrument for evaluating gain of knowledge, even considering its limitations. It could also be used for further evaluation of minor or major changes in the school curriculum in the future.

RESUMO

Tomic ER, Martins MA, Lotufo PA, Benseñor IM. Resultados de oito aplicações do Teste do Progresso na Faculdade de Medicina da Universidade de São Paulo. Clinics. 2005;60(5):389-96.

O Teste do Progresso foi introduzido na Faculdade de Medicina da Universidade de São Paulo em 2001.

OBJETIVO: (1) Testar a viabilidade da aplicação rotineira do teste; (2) verificar se o ganho de conhecimentos era progressivo e contínuo durante a graduação; (3) determinar se esse ganho de conhecimento inclui também as disciplinas do curso básico.

MÉTODOS: O teste foi aplicado duas vezes por ano entre 2001-2004. Em cada teste, calculou-se o escore médio de acertos por ano letivo usando-se ANOVA com correção de Bonferroni para múltiplas comparações.

RESULTADOS: O Teste do Progresso foi implementado como rotina entre 2001-2004. Os resultados sugerem um ganho cognitivo contínuo e progressivo ao longo da graduação ($P < 0,0001$) nos oito testes aplicados até o momento. Esse ganho seria significativo mesmo para as disciplinas do curso básico ($P < 0,05$), curso clínico ($P < 0.0001$) e internato ($P < 0.0001$). Não houve diferença de performance em função do gênero.

CONCLUSÃO: O Teste do Progresso foi implementado como rotina, sendo aplicado semestralmente. Os resultados sugerem que o ganho cognitivo parece ser contínuo e progressivo mesmo para as disciplinas do básico ao longo dos seis anos.

UNITERMOS: Teste do progresso. Teste longitudinal. Avaliação. Ensino. Graduação médica.

REFERENCES

1. Waldrop MM. Toward a unified theory of cognition. Science. 1988;241(4861):27-9.
2. Norman GR, Tugwell P, Feightner JW, Muzzin LJ, Jacoby LL. Knowledge and clinical problem-solving. Med Educ. 1985;19(5):344-56.
3. Kassirer JP. Diagnostic reasoning. Ann Intern Med. 1989;110(11):893-900.
4. Van der Vleuten CP, Verwijnen GM, Wijnen WH. Fifteen years of experience with progress testing in a problem-based learning curriculum. Medical Teacher. 1996;18(2):103-9.
5. Arnold L, Willowghby TL. The Quarterly Profile Examination. Acad Med. 1990;65(8):515-6.
6. Blake JM, Norman GR, Keane DR, Mueller CB, Cunnington J, Didyk N. Introducing progress testing in McMaster university's problem-based learning curriculum: psychometric properties and effect on learning. Acad Med. 1996;71(9):1002-7.
7. Boshuizen HP, Van der Vleuten CPM, Schmidt HG, Machiels-Bongaerts M. Measuring knowledge and clinical reasoning skills in a problem-based learning curriculum. Med Educ. 1997;31(2):115-21.
8. Muijtjens AMM, Van Mameren H, Hoogenboom RJI, Evers JLH, Van der Vleuten CPM. The effect of "don't know" option on test scores: Number-right and formula scoring compared. Med Educ. 1999;33(4):267-75.
9. Albano MG, Cavallo F, Hoogenboom R, Magni F, Majoor G, Manenti F et al. An international comparison of the knowledge levels of medical students: the Maastricht progress test. Med Educ. 1996;30(4):239-45.

10. Vosti KL, Bloch DA, Jacobs CD. The relationship of clinical knowledge to months of clinical training among medical students. *Acad Med.* 1997;72(4):1097-102.
11. Blake JM, Norman GR, Smith EK. Report card from McMaster: Student evaluation at a problem-based medical school. *Lancet.* 1995;345(8954):899-902.
12. Verhoeven BH, Verwijnen GM, Scherpbier AJJA, Holdrinet RSG, Oeseburg BM, Bulte JA, et al. An analysis of progress test results of PBL and non-PBL students. In: van der Vleuten CPM, Scherpbier AJJA, Wijnen WHFW, editors. *Progress testing: the utility of an assessment concept.* Groningen: Bas Verhoeven; 2003:125-36.
13. Schmidt HG. Innovative and conventional curricula compared: what can be said about their effects? In: Nooman ZM, Schmidt HG, Ezzat ES, editors. *Innovation in medical education. An evaluation of its present status.* New York: Springer Publishing Company; 1990. p. 1-7.
14. Borges DR, Stella RCR. Avaliação do ensino de Medicina na Universidade Federal de São Paulo. *Rev Bras Educ Med.* 1999;23(1):11-7.