

The predictive validity of the aptitude test for the performance of students starting a medical curriculum

Bernard Cerutti^a, Laurent Bernheim^b, Elisabeth van Gessel^a

^a Unit of Development and Research in Medical Education, Faculty of Medicine, University of Geneva, Switzerland

^b Vice Dean for academic Affairs, Faculty of Medicine, University of Geneva, Switzerland

Summary

INTRODUCTION: Selection of medical students varies between German- and French-speaking Swiss faculties. Geneva introduced an aptitude test in 2010, aimed at helping decision making among students. The test was compulsory: it had to be taken by those who intended to register for medical studies. But it was not selective: there was no performance threshold under which registration would have been denied.

METHODS: We followed 353 students who took the test in 2010, checked whether they confirmed their registration for medical studies and studied their performance during year 1 (selective year).

RESULTS: The correlation between the aptitude test result and the academic performance during year 1 was 0.47 ($n = 191$), and weakened to 0.38 ($n = 214$) when including repetition of year 1. The failure to pass in year 2 or success were associated with the aptitude test results ($p < 0.001$). Overall, 20% of the students succeeded after one year, 26% after a repeated year 1, and 53% failed and could not follow further medical studies.

CONCLUSION: Though there was a clear association between the aptitude test and academic performance, students did not appear to take into account when making their career decisions the ability of the test (as implemented in Geneva, that is, compulsory but not selective) to predict their future performance in the medical programme. The test was withdrawn after the 2012 session, but a number of issues regarding the medical selection procedure remain to be addressed.

Key words: student; faculty; aptitude test; selection

Introduction

Admission tests to select Medical School students are internationally widely used, and one of the main points often discussed is how well these selection tests are able to predict academic performance. They often use either cognitive elements such as prior achievement or aptitude tests, and

noncognitive measures that include interviews and personal statements; more often, however, a combination of both is used [1], in order to gain better insight into the ability of the student to follow and achieve medical studies.

Selection tests

The Medical College Admission Test (MCAT) is well known in the United States, and has shown some correlation with the United States Medical Licensing Examination (USMLE) Step 1 [2]. This written test has four main components: physical sciences, verbal reasoning writing, and biological sciences. It has been shown that when the MCAT is combined with the Grade Point Average (GPA) to improve the level of predictability, it accounts for 50% of variance in medical school at year 3, whereas the MCAT alone accounts only for 19% variance at year 1 and 15% of clinical performance. Performance at year 1 might therefore be better correlated with the GPA than with the result of a single aptitude test. However, many countries use aptitude tests for the selection of students. One of the rationales behind this organisational choice is that these tests are expected to be less biased regarding socioeconomic background and environment than previous school performance. In Australia, the Graduate Australian Medical Schools Admission Test (GAMSAT) is used to select students in various health professional programmes. It has been shown to correlate with the academic performance at years 1 and 2 [3]. A second test is used in Australia: the Medicine and Health Sciences Admission Test (UMAT), which also seems to correlate well with academic results from years 1 to 6 [3].

Some medical schools also use interviews as part of the selection process for admission. A meta-analysis of 20 studies [4] suggests that such interviews have little value to predict clinical success ($\rho = 0.17$) or academic success ($\rho = 0.06$). Noncognitive skills assessed through a personal quality assessment during medical school selection shows little correlation with Objective Structured Clinical Examinations (OSCE) [5]. Interestingly, some correlation was shown between admission structured interview scores and the average marks during later clinical years [6]. Multiple

mini-interviews have been shown to have better reliability [7]: they are a better predictor of clinical clerkship OSCE performance and clinical encounter cards than other non-cognitive measures and previous GPA.

A good review of these different selection procedures has recently been published [8], and the authors also show a strong correlation (around 0.8) between academic performance during the first three years of medical or health science studies, and both high school performance and results of the achievement test at the end of high school in Saudi Arabia.

Selection of medical students in some European countries

The United Kingdom Clinical Aptitude Test (UKCAT) has been progressively implemented since 2006. This written test has four main components that explore verbal reasoning, quantitative reasoning, abstract reasoning and decision analysis. In a fairly recent meta-analysis, the MCAT showed a correlation with academic performance in pre-clinical years [9], but the UKCAT on the other hand did not [10], and its fairness has been questioned [11]. The Irish Health Professions Admission Test (HPAT) also shows some correlation with academic achievement in medicine [12].

In other European countries such as Germany, admission to medical schools is administered by a centralised federal organisation with the universities. One important criterion for admission is the final GPA scored by the applicant in the highest secondary school diploma (Abitur), but different additional criteria are used to select students for admission, varying from one university to another (written test, structured interview, etc.). In France every student can register at a university after graduating from high school (Baccalauréat, the highest secondary school diploma), but the first year is highly selective, and every faculty has a “numerus clausus”. This first year mainly includes theoretical classes such as biophysics and biochemistry, anatomy, ethics or histology, and because of the increasing number of students, it can sometimes be reduced to a number of video-recorded courses that the students have to study at home. In Italy, an admission test is organised every year by the Ministry of Education. The test has five components: general knowledge, mathematics, physics, chemistry and biology. Every student graduating from high school (esame di maturità, the highest secondary school diploma) can register for the test. The test is administered locally: each faculty has its own “numerus clausus” and the result of the test is only valid for entry into the medical school where the examination is taken. Finally, the three Austrian public universities have so far used the same aptitude test as Switzerland, and every faculty has a “numerus clausus”.

Situation in Switzerland

Switzerland has five faculties offering a complete human medical curriculum: Basel, Bern, Geneva, Lausanne and Zurich; whereas Freiburg only offers a 3-year bachelor curriculum and Neuchatel only bachelor year 1. A selective aptitude test designed to measure the potential to follow and succeed in medical studies is used to regulate entrance of students in Basel, Bern, Freiburg and Zurich. Lausanne

and Geneva use academic performance during bachelor year 1 to select the cohort of students allowed to continue the curriculum. In 2010 Geneva introduced the same aptitude test as the German speaking faculties for a trial phase, but under different conditions: whereas the test was compulsory for any student wishing to study medicine, the score obtained by the individual had virtually no impact on the selection process. Basically, all students enrolled at the faculty of medicine, regardless of their performance score in the aptitude test. However, the third of the student cohort with the lowest scores was invited to make an appointment with a careers advisor of the local Department of Education to discuss further their interest and motivation to study medicine.

The aptitude test used in Switzerland has ten components: quantitative problems, visualisation and understanding of tubular figures, text understanding, planning and organisation, working with care and concentration, basic biology and medical reasoning, figure memorisation, fact memorisation, recognition of fragments of figures, and understanding of diagrams and tables. This test is not a “pure” aptitude test, and is designed to measure fluid intelligence (ability to think and reason abstractly and solve problems) and crystallised intelligence (ability to learn from past experiences and relevant learning, and to apply this learning to a situation).

The specific situation in Geneva allowed students to register in bachelor year 1, despite having obtained a low score in the test, and offered a unique opportunity to investigate the association between the aptitude test and academic performance without truncation, i.e., without removing the cases with low scores in the aptitude test.

Thus our main objective was to study the association between the aptitude test and the average performance scores (APS) during bachelor year 1, and the capacity to go through this selective year.

Material and methods

The APS is the average of the scores obtained in the two examinations taking place in the bachelor year 1, both of which must be passed to access bachelor year 2. One ex-

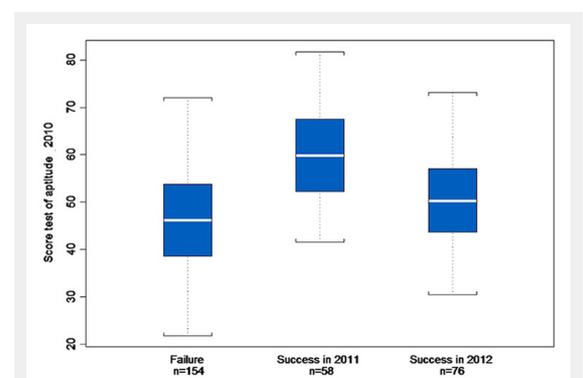


Figure 1

Boxplot of the scores obtained in the aptitude test in 2010 for the students who failed to pass in bachelor year 2 within the time allowed (two academic years), the ones who succeeded immediately (success in 2011) or after repeating year 1 (success in 2012).

am is held in January (from molecule to cell; from cell to organ; person health society; integrated clinical problems; physics, chemistry), the other in June (from organ to system; integration unit; person health society; integrated clinical problems; medical statistics). Both examinations have a duration of four hours, and include approximately 120 multiple choice questions. The APS is the mean of the two examination scores.

The APS was computed with two different approaches: the first one (APS1) took into account the results obtained only during the academic year immediately consecutive to the aptitude test; the second one (APS2) took into account the results obtained during either the first year, or the repetition year for those students who failed on their first attempt. When a student had taken an examination more than once, the most recent score was considered, since it is the score that determines whether the student is admitted to year 2.

Comparisons of continuous variables split by groups were made with analyses of variance, followed by Fisher tests. Linear regression models were used to measure the association between scores, as well as to validate the linear association hypothesis. Generalised linear model with logit link function was used to measure the association between scores and a pass or fail outcome. All tests were performed with a Type I error rate of 0.05. All the analyses were done with TIBCO Spotfire S+® 8.1 for Windows, TIBCO Software Corporation, Palo Alto, CA, USA.

Results

Among the 380 people registered for the aptitude test in 2010, 353 (92.8%) attended the test session: 51 (14.4%) did not confirm their registration for the medical curriculum in Geneva, 14 (4.0%) registered for a dental curriculum and 288 (81.6%) registered for the medical curriculum. The 51 students who finally did not confirm their registration at the Medical School achieved a lower score at the aptitude test than the others (mean score 45.6 vs 49.9; $p = 0.012$).

The test results and its components are displayed in table 1 with the correlation with both APS1 and APS2. There is a significant correlation between the aptitude test and both

APS1 and APS2, but this correlation is weaker for APS2. The aptitude test accounts for 22% of the variability of APS1 and 14% of the variability of APS2.

As with performance in the aptitude test, there is a difference between men and women regarding the APS2 (62.5 vs 67.2 for men; $n = 214$; $p = 0.022$). However, no such difference is detectable for APS1 (56.8 vs 58.7 for men; $n = 191$; $p = 0.351$).

Overall, 53% of the students (47% of the men and 58% of the women; $p = 0.275$) failed to pass the bachelor first year examinations at the end of the allowed period of two academic years. Around 20% (23% of the men and 18% of the women) succeeded after one year, and 26% after repeating bachelor year 1 (30% of the men and 24% of the women). The average scores in the aptitude test were 46.2 for those students who failed to pass in the bachelor second year, 50.8 for those who passed but who repeated the first bachelor year and 59.3 for those who passed the bachelor first year at first attempt ($p < 0.001$; fig. 1).

A logistic regression model showed a clear association between the score in the aptitude test and success during bachelor year 1 ($p < 0.001$), with no difference between genders ($p = 0.810$). With a score of 40 at the aptitude test, the probability of success without repeating is 0.06 (95% confidence interval [CI] 0.03–0.11), and 0.29 with repeating (95% CI 0.22–0.37). With a score of 50 in the aptitude test, the probability of success without repeating is 0.15 (95% CI 0.11–0.20), and 0.46 with repeating (95% CI 0.40–0.52). With a score of 60 at the aptitude test, the probability of success without repeating is 0.33 (95% CI 0.26–0.41), and 0.64 with repeating (95% CI 0.56–0.71) (fig. 2).

Discussion

This study shows a clear association between the aptitude test result and the APS during bachelor year 1, and the probability of success with or without repeating the first year. The correlation between the aptitude test and the APS1 is moderate, and weakens further when the APS2 is considered. Also, our results show an association similar to the one observed in the USA: the Geneva aptitude test accounts for 14% of the variability of data at year 1, which is similar to the 19% seen with the MCAT. The transitory and very special regulatory situation in Geneva offered the authors a unique opportunity to compare the aptitude test performance with the scores at examinations, independently of performance in the aptitude test, since the test was not selective.

Gender effect

Male students, who are a minority, tended to perform better both in the aptitude test and in the APS during bachelor year 1, although the sample size does not allow any further conclusions as to the existence of a truly different success rate ($p = 0.096$). This gender effect is well acknowledged for several selection tests [13–15]. The questions are still unanswered and remain controversial regarding the level of ability of male students who chose a medical curriculum. Is there a possible bias in the evaluation methods in their favour, or more generally the existence of a true gender dif-

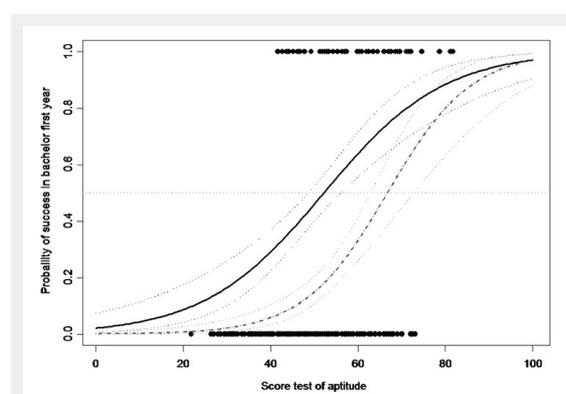


Figure 2

Probability of success during bachelor year 1 as a function of the score at the aptitude test ($n = 288$). The continuous line is the probability of success after one or two attempts and the dashed line is the probability of success without repeating year 1. The fitted lines obtained with logistic regression models are drawn with 95% confidence interval.

ference for specific aptitudes for example in quantitative reasoning?

Limitations

It was not possible to reject the hypothesis that the scores obtained by the test participants would have been different if the test had been selective; the scores in the aptitude test were on average 18% lower than at other sites in Switzerland where it was selective, this however partially confirmed the hypothesis. The absence of a huge difference in scores compared with the other sites, or some low scores considered to be statistical outliers, as well as the normal distribution of the results, may suggest that there was merely a simple shift down due to both weaker preparation and a lower level of concentration on the day of the test. If so, such a shift should not have strongly distorted the association between the test results and future academic performance. Another limitation of the study was that, although it was carried out with a relatively significant number of students, it dealt with only one cohort of students and one faculty; generalisation at this step would be hazardous. Finally, the process of decision making is much more complex than a yes or no decision based objectively on a single performance score. The study did not allow measurement of intrinsic and extrinsic elements of motivation that probably have a great influence on how the performance test results are interpreted by the students.

Effect on enrolment and decision-making

Since its introduction in 2010, the compulsory aptitude test, which has shown its ability to predict success after the first bachelor year, did not appear to be taken account by the students for their own decision to proceed or not into medical studies in Geneva. In addition, following the results of the test only two students made an appointment for a free consultation with a career advisor, among approximately 120 who had been invited to do so.

The test was rather unprofitable in its current frame, which led to several options. Firstly, its introduction as a selective test as in other Swiss or Austrian universities; the current policy of the Geneva Department of Education, which is in favour of giving everybody the opportunity to start an academic curriculum, would not have concurred with that decision. The second option was a complete removal of the test, and this was the decision made in 2012. A third option,

a selective procedure based on the test combined with multiple mini interviews [5], might be contemplated in future if the number of students in year 1 continues to increase.

Remaining issues

The success rate among the students who registered for the medical curriculum was 20.1% without repeating year 1, and 46.5% with repeating. This level of selection is lower than, or within the range of, many other European countries. In Geneva every student who has graduated from high school is given a chance to start a medical curriculum. However, the current situation raises three main concerns. First, a high proportion of students (53.5%) end up in a situation of academic failure after two years, without any possibility of starting or continuing another health-related curriculum. Second, the fact of having a large percentage of students repeating the first year creates a hold-up phenomenon that is clearly attributed to this selection barrier put at the end of bachelor year 1; this penalizes the newly enrolled students who do not benefit from the so-called "learning effect" among the students who repeat their first year. Finally, the complete shift of paradigm for the students who face a fast transition from a highly competitive and pedagogically traditional first year to a completely problem-based learning system in bachelor year 2, is not entirely satisfactory from a pedagogical point of view.

Conclusions

Medical studies are a field in which the education managers are rather lucky to have the opportunity to choose their students. A selection procedure is a highly complex multilayer and multiobjective grid, with short- and long-term goals, and every action in one layer may have consequences in all the others. In a meta-analysis conducted in 2011 [16], the authors concluded that "there is a consensus about assessment for selection for the health professions and speciality programmes, but the areas of consensus are small".

Finally, and from another point of view, it may be said that certain groups within the population are not well represented among medical school students, and that widening access may be also driven by sociopolitical concerns. Medical schools must be socially accountable and address the health concerns of communities, region or nation, but there is very little theory and empirical research on the im-

Table 1: Aptitude test scores (%) by gender and correlation with APS during Bachelor Year 1.

	All (n = 353)	Men (n = 141)	Women (n = 212)	p-value test difference between gender	Correlation with APS2 (n = 214)	Correlation with APS1 (n = 191)
Test	49.3	51.8	47.6	<0.001**	0.380**	0.469**
Quantitative problems	49.8	55.9	45.7	<0.001**	0.345**	0.464**
Tubular figures visualisation and understanding	57.1	59.2	55.6	0.088	0.139*	0.151*
Text understanding	49.0	52.4	46.8	0.005**	0.307**	0.445**
Planning and organisation	41.3	43.0	40.1	0.075	0.310**	0.340**
Working with care and concentration	55.8	61.3	52.2	<0.001**	0.198**	0.210**
Basic biology and medicine reasoning	47.8	52.7	44.5	<0.001**	0.400**	0.460**
Figure memorisation	42.5	41.3	43.3	0.309	0.131	0.143*
Facts memorisation	50.3	46.6	52.7	<.001**	0.140*	0.140
Fragments of figures recognition	51.1	52.2	50.4	0.297	0.090	0.097
Diagrams and tables understanding	48.3	54.0	44.6	<0.001**	0.304**	0.434**

* Significant result at the level 5% ** Significant result at the level 1%

pect of these considerations on selection policy [1], which is a complex issue. In the UK, the UKCAT was suspected of exacerbating rather than ameliorating the social class selection bias [14], but there were opposing views [17]. This may be true for many other selection procedures in other countries, but to our knowledge the issue has never been raised formally in Switzerland, where the number of primary care physicians in rural areas has decreased a great deal, and about 23% of the inhabitants in Switzerland are foreign (40% in Geneva). In addition to a solid reflection on selection procedures, which is on-going in many places [1, 14, 18], and in the context of a future need to train more medical doctors especially in primary care, the issue of facilitating any connexions between the different health curricula, and at different levels, is probably also a strategic choice that can help broaden the profiles of medical students while guarantying a minimum of skills and competences to be achieved with the Federal Licensing Examination in Human Medicine.

Acknowledgments: Our thanks to the three reviewers who have highly contributed to improve the structure and precision of some elements presented in the article. We are also grateful to Professor Klaus-Dieter Hänsgen from the Centre for the Development of Test and Diagnosis, University of Freiburg, Switzerland, for the fruitful collaboration regarding the implementation of the aptitude test in Geneva, as well as and the positive exchange of information and data.

Funding / potential competing interests: No financial support and no other potential conflict of interest relevant to this article was reported.

Correspondence: Bernard Cerutti, PhD, MPH, University of Geneva, Faculty of Medicine, 1 Rue Michel Servet, CH-1211 Geneva 4, Switzerland, [bernard.cerutti\[at\]unige.ch](mailto:bernard.cerutti[at]unige.ch)

References

- 1 Roberts C, Prideaux D. Selection for medical schools: re-imagining as an international discourse. *Med Educ.* 2010;44(11):1054–6.
- 2 Callahan CA, Hojat M, Veloski J, Erdmann JB, Gonnella JS. The predictive validity of three versions of the MCAT in relation to performance in medical school, residency, and licensing examinations: a longitudinal study of 36 classes of Jefferson Medical College. *Acad Med.* 2010;85(6):980–7.
- 3 Coates H. Establishing the criterion validity of the Graduate Medical School Admissions Test (GAMSAT). *Med Educ.* 2008;42(10):999–1006.
- 4 Goho J, Blackman A. The effectiveness of academic admission interviews: an exploratory meta-analysis. *Med Teach.* 2006;28(4):335–40.
- 5 Dowell J, Lumsden MA, Powis D, Munro D, Bore M, et al. Predictive validity of the personal qualities assessment for selection of medical students in Scotland. *Med Teach.* 2011;33(9):e485–8.
- 6 Mercer A, Puddey IB. Admission selection criteria as predictors of outcomes in an undergraduate medical course: a prospective study. *Med Teach.* 2011;33(12):997–1004.
- 7 Reiter HI, Eva KW, Rosenfeld J, Norman GR. Multiple mini-interviews predict clerkship and licensing examination performance. *Med Educ.* 2007;41(4):378–84.
- 8 Al Alwan I, Al Kushi M, Tamim H, Magzoub M, Elzubeir M. Health sciences and medical college preadmission criteria and prediction of in-course academic performance: a longitudinal cohort study. *Adv Health Sci Educ Theory Pract.* 2012; Jun 6. Epub ahead of print.
- 9 Donnon T, Paolucci EO, Violato C. The predictive validity of the MCAT for medical school performance and medical board licensing examinations: a meta-analysis of the published research. *Acad Med.* 2007;82(1):100–6.
- 10 Lynch B, Mackenzie R, Dowell J, Cleland J, Prescott G. Does the UKCAT predict Year 1 performance in medical school? *Med Educ.* 2009;43(12):1203–9.
- 11 Turner R, Nicholson S. Can the UK Clinical Aptitude Test (UKCAT) select suitable candidates for interview? *Med Educ.* 2011;45(10):1041–7.
- 12 Halpenny D, Cadoo K, Halpenny M, Burke J, Torreggiani WC. The Health Professions Admission Test (HPAT) score and leaving certificate results can independently predict academic performance in medical school: do we need both tests? *Ir Med J.* 2010;103(10):300–2.
- 13 Hänsgen KD, Spicher B. Aptitude test for medical studies 2012 [in German]. Zentrum für Testentwicklung und Diagnostik 2012, University of Freiburg, Switzerland.
- 14 Lambe P, Waters C, Bristow D. The UK Clinical Aptitude Test: is it a fair test for selecting medical students? *Med Teach.* 2012;34(8):e557–65.
- 15 Wright SR, Bradley PM. Has the UK Clinical Aptitude Test improved medical student selection? *Med Educ.* 2010;44(11):1069–76.
- 16 Prideaux D, Roberts C, Eva K, Centeno A, McCrorie P, et al. Assessment for selection for the health care professions and specialty training: consensus statement and recommendations from the Ottawa 2010 Conference. *Med Teach.* 2011;33(3):215–23.
- 17 Tiffin PA, Dowell JS, McLachlan JC. Widening access to UK medical education for under-represented socioeconomic groups: modelling the impact of the UKCAT in the 2009 cohort. *BMJ.* 2012;344(17):e1805.
- 18 Kaplan RM, Satterfield JM, Kington RS. Building a better physician – the case for the new MCAT. *N Engl J Med.* 2012;366(14):1265–8.

Figures (large format)

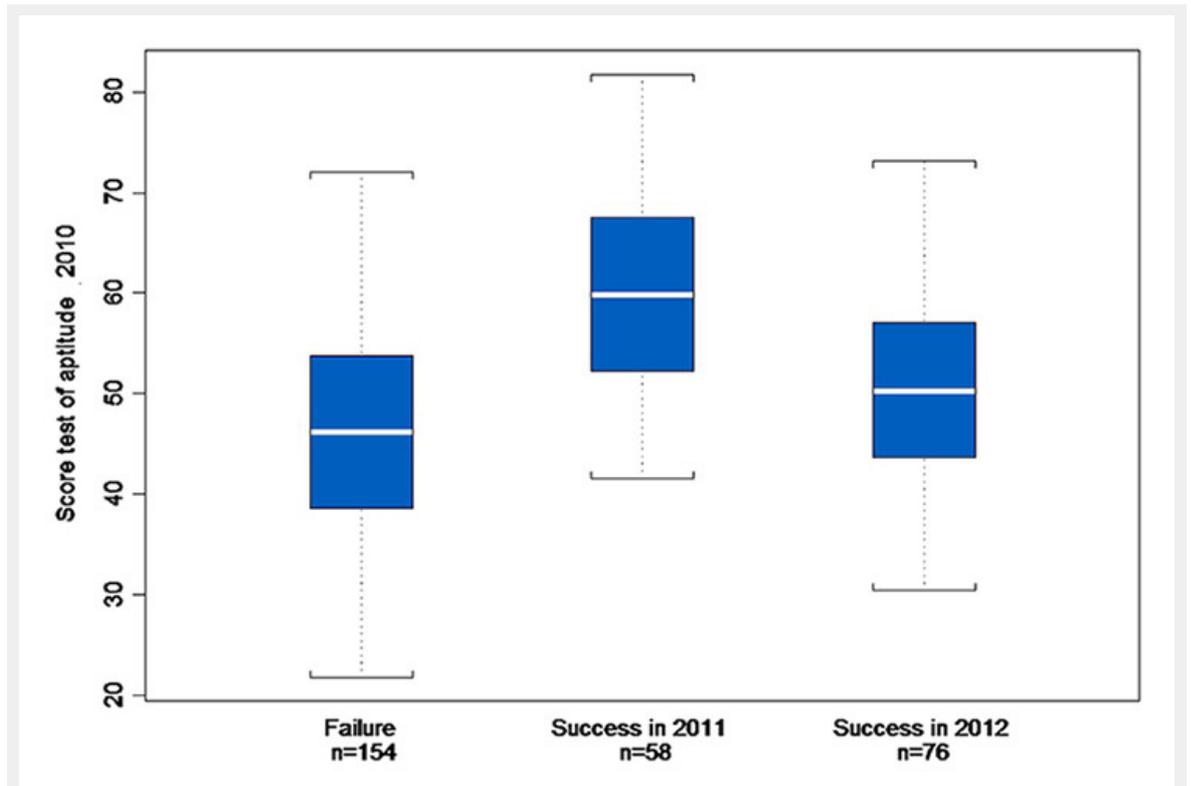


Figure 1

Boxplot of the scores obtained in the aptitude test in 2010 for the students who failed to pass in bachelor year 2 within the time allowed (two academic years), the ones who succeeded immediately (success in 2011) or after repeating year 1 (success in 2012).

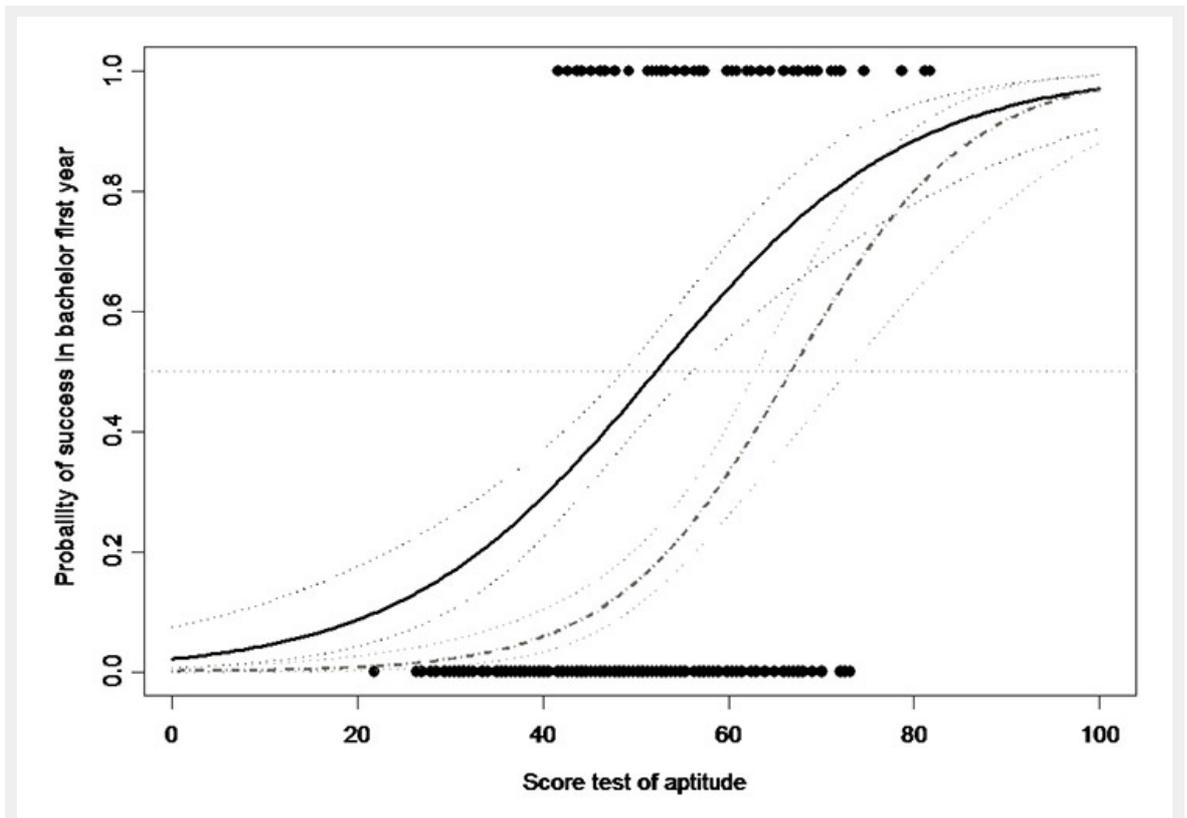


Figure 2

Probability of success during bachelor year 1 as a function of the score at the aptitude test (n = 288). The continuous line is the probability of success after one or two attempts and the dashed line is the probability of success without repeating year 1. The fitted lines obtained with logistic regression models are drawn with 95% confidence interval.