Engineering and science positioning tests in Flanders: powerful predictors for study success?

J. Vanderoost
Researcher
Tutorial Services, Faculty of Engineering Science, KU Leuven
Leuven, Belgium
E-mail: jef.vanderoost@kuleuven.be

C. Van Soom
Professor, Head of Tutorial Services Science
Tutorial Services, Faculty of Science, KU Leuven
Leuven, Belgium
E-mail: carolien.vansoom@kuleuven.be

G. Langie
Professor, Vice-Dean for Education
Faculty of Engineering Technology, KU Leuven
Leuven, Belgium
E-mail: greet.langie@kuleuven.be

J. Van den Bossche
Professor
Faculty of Engineering Technology, KU Leuven
Gent-Aalst, Belgium
E-mail: johan.vandenbossche@kuleuven.be

R. Callens
Tutor
Tutorial Services, Faculty of Engineering Science, KU Leuven
Leuven, Belgium
E-mail: riet.callens@kuleuven.be

J. Vandewalle
Professor Emeritus
Faculty of Engineering Science, KU Leuven
Leuven, Belgium
E-mail: joos.vandewalle@kuleuven.be

T. De Laet[1]
Professor, Head of Tutorial Services Engineering Science and Head of Leuven Engineering and Science Education Center (LESEC)

[1] Corresponding Author

Tutorial Services, Faculty of Engineering Science, KU Leuven
Leuven, Belgium
E-mail: tinne.delaet@kuleuven.be

Keywords: positioning test, prediction, study success

INTRODUCTION

A new battery of positioning tests for science and engineering was broadly implemented in Flanders in the summer of 2013. The goal of the non-mandatory and non-binding positioning test is to allow future students, with a clear choice for engineering or science, to position themselves with respect to the required prior knowledge and skills and to stimulate students to participate in an intervention program if necessary. For each engineering or science bachelor under study, a specific positioning test was designed. The composition of the test varies among different bachelors, as each bachelor requires specific prior knowledge and skills. For example, the positioning test for the Bachelor of Engineering Science measures the ability of future students to solve scientific problems and compares a student's mathematical skills with the required prior knowledge. The positioning test for the Bachelor of Engineering Technology and the Bachelor of Science (mathematics, physics, informatics on the one hand, chemistry, biochemistry, biology, geology and geography on the other hand) additionally tests the academic potential of the future student (such as academic literacy, logic reasoning).

For the different bachelors, the goal of this paper is to relate the study efficiency at the end of the fall semester (January 2015) to the score on the positioning test (summer of 2014). Specifically, we investigate whether the tests are able to discriminate between three groups of students: students with a high success rate in the bachelor (A-group), students in a middle group (B-group) and students with a very low success rate (C-group). We aim at providing each of these groups with the optimal feedback after the test. We want to motivate the students in the A-group to enrol in the corresponding bachelor courses and warn them that, although they have a good position on the starting grid, study success also depends on other factors. The B-group needs to be motivated and supported to improve their performance in order to succeed in the bachelor. Although the tests are non-mandatory and non-binding, the students in de C-group can be warned that enrolling in the corresponding bachelor courses is ill advised, since very few students from this group will be able to complete their bachelor successfully. We explicitly don't evaluate the quality of the positioning test using (linear) correlation statistics, since the goal of the tests is not to (linearly) correlate with the bachelor study results but to discriminate between the defined groups in order to provide optimal feedback to students participating in the tests. Therefore we introduce a methodology to compare the quality of different position tests and determine optimal thresholds for feedback based on the positioning test results.

We previously published the results and follow-up of the engineering science positioning test of 2013 [1]. In that study the impact of the positioning test on the enrolment and the study trajectory of the studies at the KU Leuven Bachelor in Engineering Science and the Bachelor in Engineering Science – Architecture were investigated. It resulted in one conclusion that is relevant for the current study. The positioning test had achieved its main goal of identifying students with low chance of study success. While passing the positioning test increases the chance on study success it does not guarantee study success. Other features such as motivation, study effort, study method, and stress-handling are important. These previously obtained results (similar to *Fig. 1)*, concerning the predictive power of the positioning test were subsequently used to convince future students to subscribe for the positioning test and to inform students about the “significance” of the obtained positioning test result. The current study not only compares this test with other positioning tests in Flanders but also reveals a scientific validation for the chosen threshold values using a new methodology.

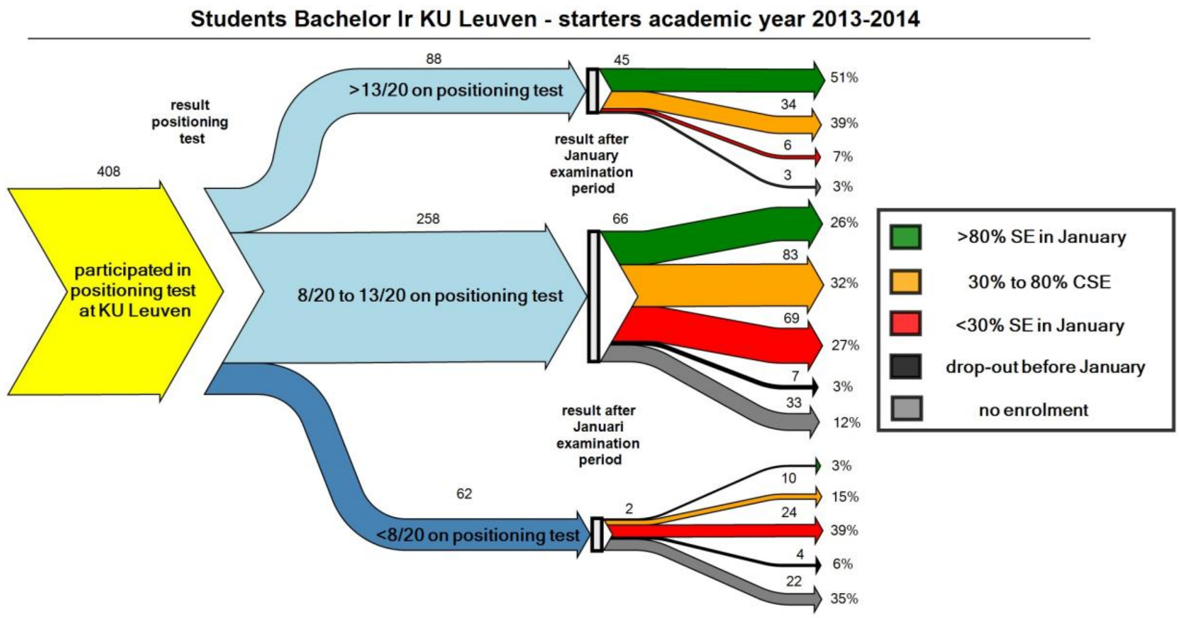

Fig. 1. Student flow diagram showing the results of the students (study efficiency SE for the January examination period) that participated in the 2013 engineering science positioning test at KU Leuven. The values alongside the arrows indicate the number of students; the percentage indicates the fraction within a specific branch.

1 AVAILABLE DATA

Since data on the full study trajectory of students is not yet available (2013 participants have not yet completed their bachelor), our analysis uses two steps to obtain a relation between the positioning test results and the final study success. First, we use the study result history of older students to determine which study efficiency (SE = percentage of obtained credits) thresholds in January (1st bachelor year) are the best classifiers to predict the number of years needed to complete the bachelor. Then we use more recent data to link the positioning test results to this January SE.

1.1 Study result history of the Bachelor of Engineering Science

In order to rate study success objectively, the history of study results was investigated for the three-year programme of the Bachelor of Engineering Science at KU Leuven going back to the 462 students that started this bachelor in academic year 2009-2010. For these students the number of years they required to obtain their bachelor degree (3, 4 or 5 years, or unsuccessful) is available and can be related to their study efficiency (SE = percentage of obtained credits) after the first examination period in January of their first bachelor year. Section 2 shows how we determined the

January SE-threshold that is best for predicting the number of years students needs to complete their bachelor.

1.2 2013 and 2014 positioning tests, study enrolment and study efficiency

Table 1 shows the different positioning tests, the number of participants, the enrolment ratio and the number of students of whom the study efficiency (SE) is available in the current study. The largest group of data was available for engineering science (IR), where students that pass the positioning test obtain a one credit exemption. This exemption is not given for the other bachelors (WIF: mathematics, physics, and informatics; CGGBB: chemistry, biology, biochemistry, geography and geology and ING: engineering technology), so by consequence they have fewer students that participated in the non-obligatory positioning tests. In the following analyses the data for SE January is pooled in one dataset if data from both academic years were available (IR, WIF, and CBBGG). For engineering technology only data following the positioning test of 2014 is available. The IR-test is solely a mathematics test, while all other positioning test include additional tests (such as academic literacy) to obtain a broader spectrum of the students' capabilities. To allow a correct comparison of different positioning tests, this study focuses primarily on the mathematics part of each test (Section 3). In Section 4 we further explore the added value of the other tests.

Table 1. Available data (number participants) on the study efficiency of participants in the different positioning tests.

position-ing test	tests	academic year	participants	% participants that enrolled	participants with SE January
IR	math	2013-2014	408	86%	350
		2014-2015	478	87%	415
WIF	math	2013-2014	90	73%	66
	academic skills	2014-2015	103	68%	69
CBBGG	math, chemistry,	2013-2014	93	69%	64
	academic skills	2014-2015	94	66%	56
ING	math, scientific reasoning, academic literacy	2014-2015	123	77%	95

2 SE VERSUS YEARS NEEDED TO OBTAIN BACHELOR

2.1 Group definitions

As mentioned in the introduction the aim of the classification is to partition the students into three groups. The A-group is defined as students who successfully complete their bachelor in three years. The B-group contains students who complete their bachelor in four years (one year study delay) en the C-group contains students who complete their bachelor in five years or don't succeed in completing their bachelor. The current section aims at classifying the students into these three groups based on the SE in January of their first bachelor year. To this end two SE-thresholds have to be chosen, one that discriminates between the A-group and the rest, and one that discriminates between the C-group and the rest dividing the classification problem into two binary prediction problems. The following analyses are based on the available data for the engineering science bachelor (see Section 1.1).

2.2 Quantifying classifier performance using ROC-curves

A Receiver Operating Characteristic or ROC-curve is a commonly used (graphical) method to quantify the performance of a binary classifier [2]. In this case it is used to quantify the ability of a SE threshold to correctly classify students into the A- versus B+C-group or C- versus A+B-group. It plots the False Positive Rate (FPR) or 1-specificity (fall-out) versus the True Positive Rate (TPR) or sensitivity (recall) corresponding to a specific threshold. The different statistics can be calculated based on values in the confusion matrix (*Table 2)* corresponding with the chosen threshold.

Table 2. Confusion matrix for C versus A+B classification using 30% SE threshold

number students: 462		***years bachelor***	
SE January		**C (5 years or no bachelor)** total true (T): 196	***A+B (3 or 4 years)*** total false (F): 266
SE <30%	**total positive (P): 156**	True Positive (TP): 137	False positive (FP): 19
SE >30%	**total negative (N): 306**	False Negative (FN): 59	True Negative (TN): 247
		True Positive Rate (TPR) = TP/T = 70%	*False Positive Rate (FPR) = FP/F = 7%*

In an ROC-curve, points on the diagonal or the line of no-discrimination are not better at discriminating between categories than a random guess. The closer the point is located to the upper-left corner (100% TPR, 0% FPR) the better the classification quality. Further on we use two different statistics that are commonly used to quantify this 'classification quality': Youden's J-statistic (vertical distance to the diagonal) and Cohen's Kappa (K-statistic) [2]. The latter takes into account the relative population sizes and accounts for the agreement occurring by chance. All of them have a maximum of one for perfect classification. The area under the ROC-curve (ROC-AUC) is a measure for the overall quality of the classifier used in the classification, irrespective of the chosen threshold. It is used here for both the quality of SE as a predictor for number of bachelor years and of the quality of the positioning test result as a predictor for the SE (Section 3). It is equal to the probability that a classifier will rank a randomly chosen 'true' instance higher than a randomly chosen 'false' instance. In the following the ROC-AUC is approximated by piecewise linear interpolation between data points.

2.3 SE January as a classifier for years of bachelor

Fig. 2 shows the ROC curves for both the A vs B+C classification and the C vs A+B classification based on the SE in January of the first bachelor year in Engineering Science. Both ROC curves correspond with very high ROC-AUC indicating that SE is a good classifier to predict the number of bachelor years (*Table 3)*. The J and K statistic can be used to determine the optimal SE threshold for each classification (maximum J and K, *Table 3)*. However, it remains important to consider the corresponding FPR and TPR values. Since the goal of the classification is to give clear feedback to students, it is important that the FPR is minimal. Therefore, we limit the FPR to maximum 10% and thereby also reduce the total positive rate PR, i.e. the fraction of the students that is marked as belonging to group A in the A vs B+C classification, and belonging to C in the C vs A+B classification. This leads to the choice of the following thresholds: SE<30% for C vs A+B and SE>80% for A vs B+C.

Table 3. ROC-AUC and optimal thresholds for classification to number of bachelor years based on SE and corresponding values for PR, FPR, TPR, J and K

classifi-cation	***classifier***	***ROC-AUC (SD)*** *[2]*	***optimal threshold***		***PR***	***FPR***	***TPR***	***J***	***K***
A vs B+C	SE Jan	0,901 (0,015)	Max. J,K	50%	52%	0,21	0,88	0,67	0,66
			FPR<10%	80%	29%	0,07	0,56	0,49	0,54
C vs A+B	SE Jan	0,893 (0,016)	Max. J,K	40%	40%	0,11	0,79	0,68	0,68
			FPR<10%	30%	35%	0,07	0,70	0,63	0,66

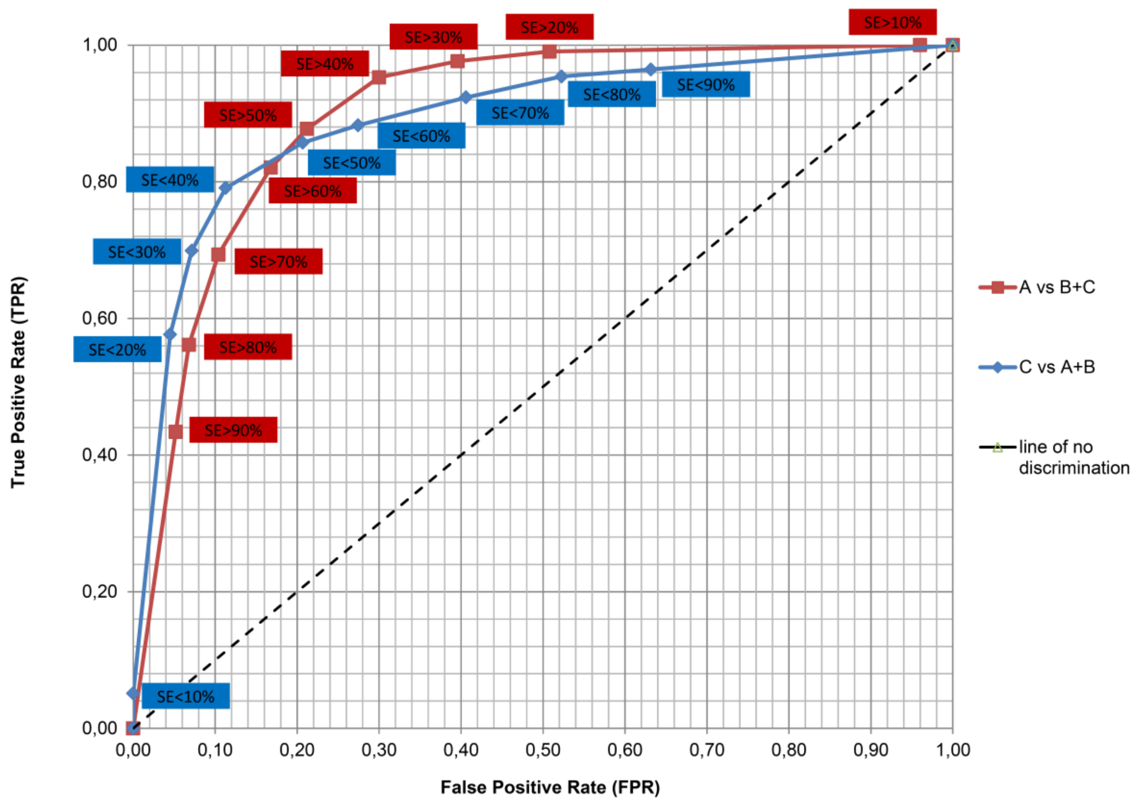

Fig. 2. ROC-curves for classification of the number of bachelor years (A vs B+C and C vs A+B) based on SE January. Markers for each step of 10% SE.

2.4 Group definitions

The SE thresholds defined in Section 2 are used to classify the students in January into three groups. The A-group is defined as students with SE>80% in January (expected to complete bachelor in 3 years). The C-group is defined as students with SE<30% in January (expected not to complete bachelor or in five years). The B-group is the middle group of students (expected to complete bachelor in four years). The current section aims at classifying the students into these three groups based on the positioning test results. Thus, two positioning test-thresholds have to be chosen, one that discriminates between the A-group and the rest, and one that discriminates between the C-group and the rest.

2.5 Mathematics positioning test results as classifier for CSE-based groups

Fig. 3 shows the ROC curves for the A vs B+C classification, a similar analysis was performed for the C vs A+B classification based on the mathematics positioning test results. The corresponding ROC-AUC can be found in *Table 4.* The smaller sample size of the WIF, CBBGG, and ING tests compared to the IR-test causes the irregular shape of the ROC-curves and affects the corresponding standard deviation (SD) of the ROC-AUC negatively. All mathematics positioning tests show a good ability to predict the SE group and are significantly different from random guessing. The IR, WIF, and CBBGG-test perform better for the C vs A+B classification, while the ING-test performs better for the A vs B+C classification, although none of these differences are significant (p>0,05*).* For each test an optimal threshold was determined by maximising the J and K statistic, while keeping the FPR below 20% for A vs B+C classification and below 10% for the C vs A+B classification (*Table 4*). This different choice of FPR-limit is related to the desired feedback. The feedback goal for

C vs A+B has a larger impact (advise not to enrol) so the FPR has to be very small, while the A vs B+C feedback is more motivational (in view of the fact that very good positioning test scores don't guarantee study success).

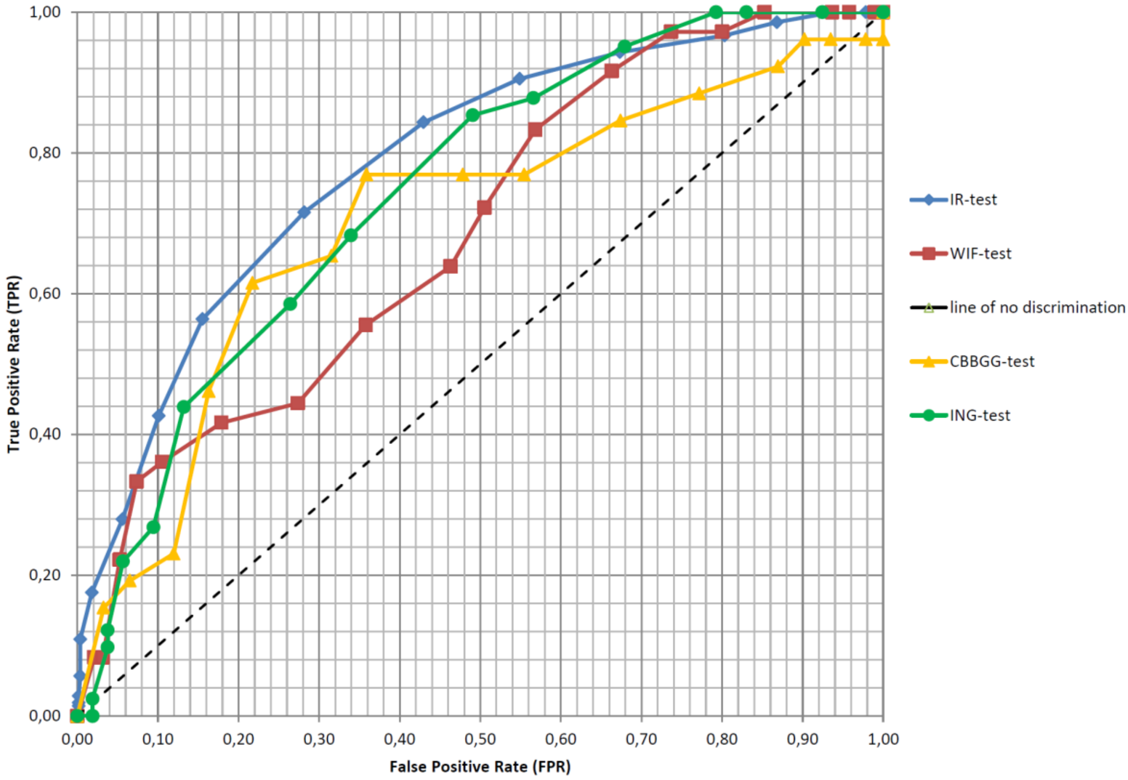

Fig. 3. ROC-curves for classification to SE groups A vs B+C in January based on the mathematics part of each positioning tests (max 20). Markers for point on 20.

Table 4. ROC-AUC and optimal threshold for classification to SE groups in January (A vs B+C and C vs A+B) based on mathematics positioning test results and corresponding values for PR, FPR, TPR, J and K

Classifi-cation	***test***	***ROC-AUC (SD)** [2]*	***optimal threshold***	***PR***	***FPR***	***TPR***	***J***	***K***
A vs B+C (SE>80%)	IR	0,757 (0,026)	>13	25%	0,15	0,43	0,28	0,30
	ING	0,707 (0,055)	>11	27%	0.13	0.44	0.31	0.32
	WIF	0.694 (0.048)	>12	27%	0.13	0.46	0.33	0.35
	CBBGG	0.724 (0.049)	>11	20%	0.13	0.32	0.19	0.21
C vs A+B (SE<30%)	IR	0,788 (0,034)	<8	12%	0,06	0,28	0,22	0,27
	ING	0,675 (0,063)	<6	12%	0,09	0,18	0,09	0,11
	WIF	0.687 (0.055)	<5	15%	0.07	0.33	0.26	0.30
	CBBGG	0.702 (0.062)	<3	9%	0.07	0.19	0.13	0.16

3 ADDED VALUE OF NON-MATHEMATICAL TESTS

Table 5 shows the ROC-AUC values for the other tests that are taken during the positioning test for the ING, WIF and CBBGG-bachelors. Overall, these additional tests are worse classifiers than the corresponding mathematics tests and are therefore less suited to classify the students. When part of a combined test they can however increase the overall classification performance. Additionally, they are able to improve the face validity of the global positioning test since students may consider the test to be more representative for their future studies then a pure mathematics test. This is, for example, likely the case for CBBGG-students and chemistry questions. Without good face validity, students will not readily accept feedback based

on the positioning test. Another reason to include a chemistry test for future CBBGG students, is to encourage students with low a chemistry test score to participate in a remedial chemistry summer course. The chemistry test is significant for the C vs A+B classification, but not to detect the best students (A vs B+C).
In the ING-test, the scientific reasoning and academic literacy do not show a significant ROC-AUC. The added value of these individual tests and the global score of these combined tests with a suitable weight will be subject of further evaluation.

Table 5. ROC-AUC for classification to SE groups in January (A vs B+C and C vs A+B) based on non-mathematical tests results. p-value for Z-test versus random classification (ROC-AUC = 0,5) shows the significance of the test.

classification	***test***	***ROC-AUC (SD)** [2]*	***p***
A vs B+C (SE>80%)	ING: scientific reasoning	0,613 (0,059)	0,063 (n.s.)
	ING: academic literacy	0.614 (0.059)	0.059 (n.s.)
	WIF: academic skills	0.696 (0.048)	<0.001
	CBBGG: academic skills	0.657 (0.052)	0.005
	CBBGG: chemistry	0.596 (0.054)	0.082 (n.s.)
C vs A+B (SE<30%)	ING: scientific reasoning	0.606 (0.065)	0.105 (n.s.)
	ING: academic literacy	0.604 (0.065)	0.108 (n.s.)
	WIF: academic skills	0.646 (0.056)	0.013
	CBBGG: academic skills	0.640 (0.064)	0.038
	CBBGG: chemistry	0.680 (0.063)	0.007

4 SUMMARY AND ACKNOWLEDGMENTS

A new battery of positioning tests for Science and Engineering was broadly implemented in Flanders in the summer of 2013. The tests allow students to position themselves with respect to the required prior knowledge and skills. The test designers aim at providing each student with an optimal feedback based on their test results. Specifically we investigated whether the tests are able to discriminate between three groups of students: students with a high success rate in the bachelor (A-group), students in a middle group (B-group) and students with a very low success rate (C-group). In this study we introduced a methodology based on ROC-curve analysis, to compare the quality of the different tests and to determine optimal threshold values for feedback. Our results show that all mathematical tests can discriminate between those three groups. The initial selected non-mathematical tests focusing on other academic skills are so far less suited for classification.

The authors acknowledge the support of the KU Leuven through the project OWP2013/27 IJkingPRO and the Leuven Engineering and Science and Education Center (LESEC).

REFERENCES

[1] Vanderoost, J., Callens, R., Vandewalle, J., De Laet, T. (2014), Engineering positioning test in Flanders: a powerful predictor for study success?, Proc. Of the 42nd Annual SEFI Conference, Birmingham, UK, 111, pp. 1-8

[2] Greiner, M., Pfeiffer, D., Smith, R.D. (2000), Principles and practical application of the receiver-operating characteristic analysis for diagnostic tests, Preventive Veterinary Medicine, Vol. 45, pp. 23-41.