Predicting performance of 1st year engineering students and the importance of assessment tools therein

This item was submitted to Loughborough University's Institutional Repository by the/an author.

Citation: LEE, S. ... et al. 2008. Predicting performance of 1st year engineering students and the importance of assessment tools therein. Engineering Education, 3 (1), pp. 44-51

Additional Information:

- This article was published in the journal, Engineering Education.

Metadata Record: [https://dspace.lboro.ac.uk/2134/8784](https://dspace.lboro.ac.uk/2134/8784)

Version: Published

Publisher: Higher Education Academy © The authors

Please cite the published version.
This item was submitted to Loughborough’s Institutional Repository (https://dspace.lboro.ac.uk/) by the author and is made available under the following Creative Commons Licence conditions.

For the full text of this licence, please go to:
http://creativecommons.org/licenses/by-nc-nd/2.5/
Predicting performance of first year engineering students and the importance of assessment tools therein

Stephen Lee, Martin Harrison, Godfrey Pell and Carol Robinson

Abstract
In recent years, the increase in the number of people entering university has contributed to a greater variability in the background of those beginning programmes. Consequently, it has become even more important to understand a student’s prior knowledge of a given subject. Two main reasons for this are to produce a suitable first year curriculum and to ascertain whether a student would benefit from additional support. Therefore, in order that any necessary steps can be taken to improve a student’s performance, the ultimate goal would be the ability to predict future performance.

A continuing change in students’ prior mathematics (and mechanics) knowledge is being seen in engineering, a subject that requires a significant amount of mathematics knowledge. This paper describes statistical regression models used for predicting students’ first year performance. Results from these models highlight that a mathematics diagnostic test is not only useful for gaining information on a student’s prior knowledge but is also one of the best predictors of future performance. In the models, it was also found that students’ marks could be improved by seeking help in the university’s mathematics learning support centre. Tools and methodologies (e.g. surveys and diagnostic tests) suitable for obtaining data used in the regression models are also discussed.

1. Introduction
All UK institutions offering engineering courses expect students to have studied mathematics, physics or other numerate subjects as a pre-requisite. It is important to have an understanding of the prior knowledge that engineering students actually have upon entry to university. Currently, departments have information on students such as their total A-level points score and their individual A-level grades. However, as found by Adamson and Clifford (2002): ‘It is clear that student performance in a university environment cannot be reliably predicted from performance indicators gained in school examinations.’

This is in agreement with Todd (2001) who concluded that: ‘As a method of selecting students for admission to our courses A-level grades give a reliable method of establishing a threshold though they are an unreliable indicator of subsequent performance.’

This was one of the reasons why mathematics diagnostic testing has become widespread in universities since the end of the last decade, as reported upon in Measuring the Mathematics Problem (Hawkes and Savage, 2000). Diagnostic testing may provide a more detailed insight into what students do or don’t know, but is such a test actually a better predictor for future performance than examination results from school?

A brief introduction to the issue of the changing prior knowledge of engineering students upon entry to university is given below. This is detailed through discussion of recent changes in pre-university qualifications and the associated issues. The research questions that motivated this work and the accompanying methodology used to answer them are then detailed. Thereafter follows the main focus of the paper which involves a discussion of linear regression models created to predict future performance. This involves consideration of 14 variables (including students’ gender and their total A-level points score) to establish which are statistically significant predictors of future performance. In the models created it can be seen that students’ results in a mathematics diagnostic test and, where available, additional help obtained at the university’s mathematics learning support centre are two statistically significant factors when predicting future performance. Comments on the reliability of the
All this did nothing to help ‘the mathematics problem’ which describes the lack of mathematical ability of students entering numerate degrees and which has been of concern during the past decade. Several major reports on this have been published, including Tackling the Mathematics Problem (London Mathematical Society, Institute of Mathematics and its Applications and Royal Statistical Society, 1995) and Measuring the Mathematics Problem (Hawkes and Savage, 2000). In recent years an associated issue labelled ‘the mechanics problem’, which centres on students’ (lack of) knowledge of mechanics on entering engineering degrees, has become evident, see Robinson et al. (2005). It should be noted that the majority of engineering students at university study compulsory modules in mathematics as well as mechanics. At Loughborough University one-sixth of an engineering student’s first year programme consists of a mathematics module and a further one-sixth of their first year programme may consist of a mechanics module.

3. Research questions and methodology

Academics are evidently interested in how students perform once at university. Being able to predict how students may perform in the future is useful for many reasons, in particular to identify those students most likely to fail. To try to address this a primary research question was posed and suitable methodology for answering it given.

The primary research question was:
What factors are significant predictors of an engineering student’s
a) overall first year university performance (including mathematics and mechanics modules) and
b) grades in their first year university mechanics module?

The research question was answered by collecting relevant data on individual engineering students so that regression models could be created. Thus, the factors which did and did not affect both performance in a specific first year university module (mechanics) and overall first year university performance were identified. In order to collect relevant data various methodologies were used (e.g. questionnaires and diagnostic testing) and these will be discussed later.
4. Can first year performance be predicted?

In order to create statistical regression models to predict student performance a large amount of data (on 133 students studying mechanical engineering courses at Loughborough University) was collected. Mechanical engineering students are a particularly relevant group, as mechanics is important to their programme and there is a good range of abilities within the group. Data collected on these students included:

- Mathematics diagnostic test mark
- Mechanics diagnostic test mark
- Mathematics A-level grade
- Gender
- Whether the student studied A-level further mathematics
- Number of mechanics modules studied in A-level mathematics
- Number of statistics modules studied in A-level mathematics
- Number of discrete modules studied in A-level mathematics
- Total A-level points score
- Whether the student studied with exam board AQA (Assessment and Qualifications Alliance)
- Whether the student studied with exam board OCR (Oxford, Cambridge and the Royal Society of Arts Examinations)
- Whether the student studied with Welsh/Northern Irish exam boards
- Whether the student was overseas or home/EU
- Whether the student visited the Mathematics Learning Support Centre (MLSC) in their first year of study.

Data such as student gender and total A-level points score was readily available from university records. Other information needed to be collected by administering diagnostic tests or questionnaires. Discussion of these methods follows in section 5. Firstly (in section 4.1), linear regression models, using the stepwise method, are discussed.

It should be noted that data could not be collated for all 133 students for all the variables mentioned above. For example, only 127 students undertook the mathematics diagnostic test and only 124 undertook the mechanics test. The six students who did not complete the mathematics test may not be the same as the nine that did not complete the mechanics test. This issue will be discussed further in section 4.3.

Once the data had been collated, a linear regression model was produced using the statistical package SPSS. This is a mathematical equation in which a variable of interest, for example overall first year performance of a student, is predicted by relating it to other variables. The numerical component of each of these other variables in the model is called their coefficient.

4.1 Regression model for overall first year university performance

Considering the overall percentage mark, \( y_i \), of mechanical engineering students in their first year, with respect to the 14 variables stated earlier, the following linear regression model was produced for the 66 students with full data sets:

\[
y_i = 0.353a_i - 5.321b_i + 7.781c_i + 35.886
\]

The variables \( a_i \), \( b_i \), and \( c_i \) are those indicated in Table 1. Also in the table are the possible values which each of the variables could take. In Table 1 there are some other standard statistical measures, the standard error and the t value, which are both inherently connected to perhaps the most important measure, the level of significance. Variables in the regression models produced meet a certain level of significance. This is the reason why all 14 variables are not present in the model above. Here 0.1 was chosen as the level of significance although, as can be seen from column six, all of the variables are statistically significant (\( p < 0.05 \)).

An R\(^2\) value of 0.392 was obtained for this model. This indicates that 39% of the variation in overall first year results could be attributed to the variables \( a_i \), \( b_i \), and \( c_i \).

When considering this model it is important to note what size each variable could take, seen in the final column, and what effect they could have on the model. It can be seen that the variables \( b_i \) and \( c_i \) can only take a small number of different values (\( b_i = 0, 1, 2, 3 \) and \( c_i = 0, 1 \)), whereas \( a_i \) could take a larger number of (discrete) values between 0 and 100, namely \( [0, 2.5, 5, …95, 97.5, 100] \), given that there were 40 questions in the test. However, as each of these have a different coefficient in the model, variables can only have a certain effect on the overall model. For example, variables \( b_i \) and
c1 multiplied by their coefficient can only take a small number of different values between -16.0 (3 \times -5.321) and 7.8 (1 \times 7.781) in the model, whereas a1 can have an effect of up to 35.3 (100 \times 0.353).

Note the value of the coefficient 7.781. This shows the positive effect of almost one grade boundary of students visiting the MLSC. Visiting the MLSC is seen to be useful not only for the less well-prepared students but also for the average and good students. The MLSC is a resource centre that students can visit at any time (between 9 am and 5 pm) to obtain assistance and guidance on mathematics. A member of the school of mathematics staff is always on duty in the MLSC when it is open. At Loughborough the MLSC is well established and has been in operation for ten years. Furthermore, a second classroom in a different location was opened for the 2006-07 academic year to give provision for the increasing number of students from a diverse range of degrees (e.g. social sciences) seeking assistance with mathematics.

From the value of the coefficient b1 it is evident that, in predicting overall first year performance, the model indicates that the study of statistics modules in A-level mathematics has a negative effect. A potential reason for this could be that studying more statistics modules leaves less opportunity to study mechanics modules which, for mechanical engineering students, is likely to have a detrimental effect.

In reviewing this model it is interesting to note which variables (of the original 14) do not appear to be significant. In particular, the usual way of selecting students for university courses by their total A-level points score was not a significant variable in the model. Perhaps most noteworthy is that the mathematics diagnostic test was a significant predictor of overall first year performance, more so than even the mechanics diagnostic test.

### 4.2 Regression model for performance in a first year university mechanics module

A second model was created to specifically consider what factors affected students’ performance in the first year university mechanics module. The same 14 variables were considered when creating the model for first year mechanical engineering students’ performance in their first year mechanics module, y2. The regression model can be seen below and the respective variables in Table 2.

An R² value of 0.476 was obtained for this model.

\[
y_2 = 0.518a_1 - 6.785b_1 + 8.949c_1 + 22.497
\]

Again it can be seen that a dominant feature of the model was the mathematics diagnostic test result, which can have an effect of between 0 and 51.8 (100 \times 0.518) in the model. In this model the number of statistics modules studied in A-level mathematics again had a negative effect. The positive effect of visiting the MLSC can again be seen, as well as a lack of significance of students’ total A-level points score. It was very interesting to observe that the same variables emerged as being significant in both models (i.e. for overall first year performance and for performance in a specific (mechanics) module). However, this may not have been surprising given the fact that the first year mechanics module is in fact a subset of the overall first year performance (i.e. students’ first year mechanics module marks contribute one-sixth of the total marks for the overall first year performance). The data was finally checked for interactions between the variables in the models using ANOVA, but none were found. Furthermore, none of the

<table>
<thead>
<tr>
<th>Variable</th>
<th>Coefficient</th>
<th>Standard error</th>
<th>t</th>
<th>Significance</th>
<th>Possible values</th>
</tr>
</thead>
<tbody>
<tr>
<td>Constant</td>
<td>35.886</td>
<td>6.007</td>
<td>5.974</td>
<td>0.00</td>
<td></td>
</tr>
<tr>
<td>a1</td>
<td>.353</td>
<td>.074</td>
<td>4.752</td>
<td>.000</td>
<td>0 - 100</td>
</tr>
<tr>
<td>b1</td>
<td>-5.321</td>
<td>1.517</td>
<td>-3.508</td>
<td>.001</td>
<td>0,1,2,3</td>
</tr>
<tr>
<td>c1</td>
<td>7.781</td>
<td>2.749</td>
<td>2.830</td>
<td>.006</td>
<td>0 - No, 1 - Yes</td>
</tr>
</tbody>
</table>
excluded variables (those that were not in the models) had a high correlation with the model predictors.

4.3 Reliability of the regression models constructed

In this section, two regression models for performance of mechanical engineering students in their first year university mechanics module and their overall first year engineering programme have been presented. However, it is appropriate to comment on the reliability of the models and establish if they could be extended to other (larger) groups of students.

Firstly, as discussed at the beginning of section 4, data for all variables could not be collated for all 133 students. Consequently, data for all variables (i.e. complete data sets) were obtained for only 66 students. Therefore, the models described previously were constructed using a stepwise method on the initial 14 variables for 66 students. Subsequently, the variables that were shown to be significant were taken and regression modules were created using only these (three) variables for all students that had complete data for them. 107 students were used in the analysis and the regression models found for overall performance \( y_1 \) and performance in the first year mechanics module \( y_2 \) were:

\[
\begin{align*}
y_1 &= 0.236a_1 - 2.274b_1 + 4.794c_1 + 40.611, \quad R^2 = 0.185 \\
y_2 &= 0.412a_1 - 3.334b_2 + 7.416c_2 + 24.940, \quad R^2 = 0.316
\end{align*}
\]

Here the variable ‘a’ represents the mathematics diagnostic test result; ‘b’ the number of statistics modules studied in A-level mathematics and ‘c’ whether a student visited the MLSC or not.

When the models were extended to include students with complete data sets for only the three significant variables (and not all 14 variables), lower \( R^2 \) values were found. For overall first year performance the \( R^2 \) value was 0.392 for the 66 students with complete data sets but 0.185 for the 107 students who had data on the three significant variables. Similarly, for performance in the first year mechanics module the \( R^2 \) value was 0.476 for the 66 students with complete data sets but 0.316 for the 107 students who had data on the three significant variables. This, along with the change in the size of the coefficients would indicate that the fit of the model(s) was not very robust.

Given that there are so many factors that could affect a student’s performance but which it has not been possible to build into the models (e.g. personal factors such as financial or accommodation issues, as described in Murdoch-Eaton et al., 2007), then values for \( R^2 \), like those found, can be seen to indicate that the significant variables are of importance. When models were created considering only data from the three significant variables \( R^2 \) values of 0.185 and 0.316 were obtained.

This indicates that 19% of the variation in a student’s overall first year result, from the many possible factors, could be attributed to the three variables detailed. Similarly, 32% of the variation in a student’s first year mechanics module result, from the many possible factors, could be attributed to the same three variables. This suggests that the three variables found to be significant in the models are certainly important.

5. Further discussion of significant predictors

In the previous section linear regression models were created to predict students’ performance. In these models factors such as a student’s mathematics diagnostic test mark, whether

| Table 2. Regression model of mechanical engineers’ first year performance in their mechanics module |
|---|---|---|---|---|---|---|
| Variable | Coefficient | Standard error | t | Significance | Possible values |
| Constant | 22.497 | 6.883 | 3.289 | 0.02 |
| \( a_1 \) | 0.518 | 0.085 | 6.088 | .000 | 0 - 100 |
| \( b_1 \) | -6.785 | 1.738 | -3.904 | .000 | 0,1,2,3 |
| \( c_1 \) | 8.949 | 3.150 | 2.841 | .006 | 0 - No, 1 - Yes |

<table>
<thead>
<tr>
<th>Variable</th>
<th>Coefficient</th>
<th>Standard error</th>
<th>t</th>
<th>Significance</th>
<th>Possible values</th>
</tr>
</thead>
<tbody>
<tr>
<td>( a_2 )</td>
<td>2.274</td>
<td>0.85</td>
<td>2.668</td>
<td>.009</td>
<td>0 - 100</td>
</tr>
<tr>
<td>( b_2 )</td>
<td>-3.334</td>
<td>1.738</td>
<td>-1.904</td>
<td>.059</td>
<td>0,1,2,3</td>
</tr>
<tr>
<td>( c_2 )</td>
<td>7.416</td>
<td>3.150</td>
<td>2.341</td>
<td>.021</td>
<td>0 - No, 1 - Yes</td>
</tr>
</tbody>
</table>
a student visited the MLSC and how many statistics modules they had studied in A-level mathematics were all found to be significant. However, detailed data on these significant factors (along with some of the non-significant factors) was not readily available and thus needed to be obtained. Here we give details of how such data was collected, as well as giving further consideration to the three factors that were present in both regression models.

5.1 Assessment tools used for obtaining additional data

Firstly, a simple questionnaire was constructed which was incorporated into a mechanics diagnostic test (written by the authors) to establish how many applied (i.e. mechanics, statistics or discrete) modules engineering students had studied at school. Note that this relied on the students having studied A-level mathematics, although it was found that approximately 90% of students surveyed had done so. In addition, it ascertained which examination board students had studied with. The questionnaire itself was not only created so that data could be used for a predictor model but also to gain an understanding of which applied modules students had studied before coming to university. Consequently, it was administered at Loughborough University in the academic year 2003-04 and extended to include three universities in the academic year 2004-05.

Multiple-choice diagnostic mathematics tests have been used with engineering students at Loughborough University for ten years. They have primarily been used to determine which students are in need of additional mathematics support, particularly during their first year. For the academic year 2004-05 engineering students also completed a mechanics diagnostic test. This test, developed by the authors, was a multiple-choice test marked electronically by an Optical Mark Reader (which was the principal reason why the questionnaire mentioned earlier was incorporated into it). Its primary purpose was to establish incoming students’ knowledge of mechanics, given the growing concern over such an issue, see Kitchen et al. (1997) and Mustoe (2004). Engineering students’ results for the mechanics diagnostic test were reported upon in Lee et al. (2005).

Data such as a student’s gender, A-level points score and A-level mathematics grade, along with their overall first year result and result in their first year mechanics module, were obtained from the respective departments. Finally, data on whether students had visited the MLSC was obtained from records held by the Mathematics Education Centre. It should be noted that students are required to ‘swipe in’ with their ID card upon entering the MLSC and thus an electronic record of who has visited the MLSC is kept.

Thus, it can be seen that various methods were used to obtain the data required to calculate the regression models. In some instances a considerable amount of time and effort was needed to create such methodologies. For example, it is not a trivial task to create a good quality diagnostic test or questionnaire. In addition, implementing, collating and analysing data from such methodologies also incurs significant time costs. Deciding whether such time and effort is justified is obviously of importance.

5.2 Significant factors

The significant factors in both regression models were: students’ mathematics diagnostic test result; whether they had visited the MLSC in their first year of study and the number of statistics modules they had studied in A-level mathematics.

In many universities mathematics diagnostic tests are already in place, as reported upon in Hawkes and Savage (2000). However, other diagnostic tests may not be good predictors and certainly could not be used with our model, though a copy of our test is available upon request. In addition, as reported by Perkin and Croft (2004), there is an ever-increasing number of mathematics support centres, in various forms, in universities in the UK. Again, data could be collected on whether a student had visited a particular mathematics support centre. However, a university would need to develop its own predictor model to establish if whether a student visits their support centre is a significant predictor. From September 2006 data on which modules students had studied in a particular A-level (i.e. the number of statistics modules in A-level mathematics) became available to universities through a student’s UCAS application. Thus, this is the only factor that can be readily used by all.

Therefore, it can be seen that another university could not just use the exact regression models
created for our students. Other universities could create their own regression models, but would this be worthwhile? Ultimately, it depends upon the motive for predicting students’ future performance. If it is to identify students who may be in need of assistance then administering a mathematics diagnostic test and then offering and monitoring subsequent support can be very beneficial, as discussed by Robinson and Croft (2003). They comment: ‘Early indications are that the diagnostic test is a useful vehicle for identifying students in need of extra support…’. This arose as a result of reviewing the 1000 mathematics diagnostic tests that are administered annually at Loughborough University. Otherwise, those interested in creating such regression models should keep in mind the considerable amount of time and effort that would be required to produce them.

6. Concluding remarks
In this paper consideration has been given to creating models for predicting engineering students’ first year university performance. The methods used to collect the appropriate data for the regression models have also been detailed. In both models three factors emerged as being significant. These included students’ mathematics diagnostic test results, whether they had visited the MLSC in their first year of study and the number of statistics modules they had studied in A-level mathematics.

Models using all 14 variables, created for 66 mechanical engineering students, were found to have higher R² values than for those created for 107 students (in which only the three significant factors were considered). This highlighted that the fit of the model(s) was not very robust. However, there were obviously a large number of factors which could have an effect on first year performance but which were not included when creating the predictor models (i.e. factors such as personal issues). Consequently, the R² values for the models created from the group of 107 students indicated that the significant variables are of importance.

Although the regression models created are specific to the group of students here, the idea of creating such models elsewhere is indeed a distinct possibility. However the time and effort needed to produce them should not be underestimated, specifically the need to implement methodologies such as questionnaires and diagnostic tests. Nevertheless, there will always be a large number of other factors that cannot be built into such models. Such remarks were also made in a similar study into first year performance by Hunt et al. (1995):

It is impossible to separate out in a quantitative way the effects of preparation, motivation and ability of the student, and the course provision of the University on the success of individual students. This is because there is a great deal of feedback between the different factors. There are many other potential factors that may be involved. However, it has been possible to point out some interesting relationships and give some warnings of current and potential problems.

An ever-changing student intake with a more diverse background means that there is the likelihood that there could be an increase in students needing support, especially in their first year. Hence, using methodologies such as diagnostic tests and then offering appropriate support has become a valuable strategy to implement with students on many courses. However, an additional approach, to cite one of many possibilities, could be to include proactive mentoring of students using learning contracts (where students’ weaknesses are identified and they agree to review their skills in the given area/s). Finally, it would be worthwhile to consider some programme development (e.g. of specific first year modules), particularly in engineering, following on from diagnosing incoming students’ knowledge.

References


About the authors

Stephen Lee BSc (Hons), PhD., AMIMA, (principal author), Professional Officer, Mathematics in Education and Industry (MEI), Trowbridge, Wiltshire, BA12 0XG. Email: stephen.lee@mei.org.uk

Martin C. Harrison BSc (Hons), MSc, PhD., CMath, MIMA, Associate Dean (Teaching), Faculty of Science, Loughborough University, Leicestershire, LE11 3TU. Email: m.c.harrison@lboro.ac.uk

Godfrey Pell BEng, MSc. C.Stat, FRSS, Statistician, School of Education and the Medical Education Unit, University of Leeds, Leeds, LS2 9JT

Carol L. Robinson BSc (Hons), PhD., CMath, FIMA, Assistant Director (Teaching and Learning), Mathematics Education Centre, Loughborough University, Leicestershire, LE11 3TU. Email: c.l.robinson@lboro.ac.uk