Seeing the Future

What CAO entrant information can tell us about students' ability to progress in college

Kevin Kelly,
School of Engineering,
TCD
24th January 2013

What I'll talk about

• The ATTRACT project
• Context
• University entrance (TCD Engineering)
• Data
• Analysis
• Results
• Implications
• Conclusions and Future Work
The ATTRACT Project

3 Year EU Project (www.attractproject.org) seeking to establish best practice in engineering education under the headings of:

- Perception
- Barriers
- Outreach
- Retention

Perception

What is the image and understanding of engineering amongst potential students?

- Media Analysis
  - Generally positive
- Employment Prospects
  - Good to excellent
- Student Surveys
  - Poor understanding of what engineers do
  - Well paid
  - Good job prospects
  - High social contribution
  - Most difficult thing to study
Outreach
What is being done and what is effective in enticing students to study engineering?

- Survey of current students
- Survey of secondary school students
  - Significant gender differences
    - Boys like the technology/output
    - Girls like (but are largely ignorant of) the social contribution
  - Parents and teachers are key influencers – need to reach them, not just students
  - Engineering is about doing – students who get a sense of this are more strongly influenced. Active learning?
  - Institutional reputation is largest factor in student choice

Retention
How do we ensure we retain as many students as we should?

- Definition – far from trivial!
- First year (and first 6 weeks) are key
- Significant mismatch of expectations in many cases
- Academic and social integration are key
- Co-curriculum can play a significant role
- Successful action depends on having processes and systems to manage:
  - Expectations
  - Feedback
  - Support
  - Involvement
Barriers
What barriers inhibit/block potential engineering students?
- Socio-economic
  - Engineering programmes have particular problems attracting LSE status students
- Structural
  - How (secondary) school structure and practice can limit choice
  - Evidence that students choose subjects based on perceived easiness of obtaining CAO points
  - Schools may not provide (or encourage) ‘engineering-relevant’ subjects
- Entrance requirements
  - Purpose
  - Appropriateness & Effectiveness

Political Context
Widespread agreement that the current system, while fair, is deeply flawed
- encourages lower order (cf Bloom) learning at the expense of higher level
- is stressful for LC candidates
- perpetuates socio-economic inequity
- creates a mindset that is unsuited to either higher education or the workplace
Agreement that entry system needs reformation
- Hyland report
- Minister Quinn
- TCD Admissions Feasibility Study
National Educational Context

- CAO
  - ~400K course applications at level 8 (37K in engineering)
  - ~67K applicants (6.5K)
  - ~45K offers (4.5K)
  - ~30K places accepted (3K)
- Funding
  - €1.1Bn for higher education
  - €6-€10K per UG student per annum
- Retention
  - ~11% of all university students don’t make it to second year
  - ~20% of TCD engineering students repeat/drop out from first year

International Educational Context

"the number of students reaching level 5 or 6 in mathematics and science will be particularly important for countries wishing to create a pool of workers able to advance the frontiers of scientific and technological knowledge in the future and compete in the global economy"

- OECD Report – PISA 2009 Results Vol. 1
## International Educational Context

### Percentages of Students at Level 5(6) in Maths and Science

<table>
<thead>
<tr>
<th>Country</th>
<th>Maths</th>
<th>Science</th>
</tr>
</thead>
<tbody>
<tr>
<td>Shanghai</td>
<td>24(26)</td>
<td>20(4)</td>
</tr>
<tr>
<td>Finland</td>
<td>16(6)</td>
<td>15(4)</td>
</tr>
<tr>
<td>OECD average</td>
<td>9(3)</td>
<td>7(2)</td>
</tr>
<tr>
<td>Ireland</td>
<td>5(1)</td>
<td>8(1)</td>
</tr>
</tbody>
</table>

### Time spent on Maths

![Graph showing time spent on Maths](image-url)

- Ireland
- Finland
- Portugal
- Sweden
- Italy
### Time spent on Physics

#### Number of Hours

<table>
<thead>
<tr>
<th>Year</th>
<th>Ireland</th>
<th>Finland</th>
<th>Portugal</th>
<th>Sweden</th>
<th>Italy</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>5</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>6</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>7</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>8</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>9</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>10</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>11</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>12</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>13</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>14</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>15</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>16</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>17</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>18</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

#### Student's Age

- **1** to **18**

---

### Pisa Scores in Maths and Science, with GDP per capita

- **Ireland**
- **Finland**
- **USA**

- **GDP per capita**

### Trends

- Positive correlation between educational investment and Pisa scores.
- Finland and Ireland show high educational investment.
- USA has a moderate relationship.

---

**International Educational Context**

**Time spent on Physics**

**Pisa Scores in Maths and Science, with GDP per capita**
So what are the barriers?

<table>
<thead>
<tr>
<th>TCD</th>
<th>Other Universities/HEI</th>
</tr>
</thead>
<tbody>
<tr>
<td>A C3 (i.e. 55%+) in Higher</td>
<td>~80% share this</td>
</tr>
<tr>
<td>Maths is required</td>
<td>~10% require a D3 in a lab science</td>
</tr>
</tbody>
</table>

What is the stated purpose? Are they effective and appropriate?
- Demarcation of teaching responsibility
- Identification of students with maximal aptitude

Data Used

Data for all LC entrants to TCD in the 2000-2009 period
- CAO points total
- CAO scores in all subjects
- Living at home (is term address different)
- Gender
- Degree programme entered
- Year of entry
- Pass/fail data from summer, supplemental and special exams in first year

Approximately 20,000 students, of whom 1850 are in engineering programmes
Analysis – Missing Data

No student will have taken all subjects, so we need to deal with that

- Restrict analysis to select subjects and only those students who’ve taken those subjects
- Use a ‘dummy variable’ to indicate whether students have taken any particular subject

Analysis - Data Formulation

- Inputs
  - Whether a student took a particular subject (binary)
  - CAO grade achieved in each subject (discrete, 0-100)
  - Degree (one of two available) programme chosen (binary)
  - Gender (binary)
  - Year of entry (ordinal, 1-10)
  - CAO score (discrete, 0-600)
  - CAO average (discrete, 0-600) – include all subjects
  - Living at home (binary)

- Outputs
  - Had to take repeat exam sessions (binary)
  - Progressed to 2nd year (binary)
We are trying to predict a binary output, using a range of variable types as inputs, so we....

1. Combine variables to produce a single number – higher indicating a greater likelihood of progression
2. Use a 'link function' to map this number to the range 0-1 (i.e. our output is the probability that someone progresses)

1. Use linear regression
2. Use a logistic function

\[
\begin{align*}
\text{Input Variables} & \rightarrow \text{Weighted Sum} & \rightarrow \text{Logistic Function} & \rightarrow \text{Probability} \\
\begin{bmatrix} \text{[0 / 1]} \\
\text{[0 - 100]} \end{bmatrix} & \rightarrow \begin{bmatrix} -\infty, +\infty \end{bmatrix} & \rightarrow \frac{1}{1 + e^{-\gamma}} & \rightarrow \begin{bmatrix} 0 - 1 \end{bmatrix}
\end{align*}
\]

The result of the modelling process will be a series of 'weights' — indicating the relative influence of the variables. However, the range & properties of the variables are also important.

- For example a CAO grade is 0 (0-39%), 45-90(40-85%), or 100 (90-100%)
- Not everyone takes every subject, so we need to distinguish between a 'null' (didn’t take the subject) and a zero (got between 0 and 39%)

By looking at the combined effect of the indicator variable and the CAO score for any subject we will have a negative effect for the indicator variable, but offset by the CAO grade received.
Presentation of Model results

Exam Mark versus Effect Size

Effect Size per Subject

'Critical Mark' (i.e. that mark above which the student benefits from taking the subject) is 65%
Only statistically significant results (p < 0.05) are shown:

- Mathematics has largest (positive) effect – but all students must have C3+, so there is confounding here with any intercept
- Largest influence is Building Construction, but only a positive benefit for those getting an A2+
- Of 'mainstream' subjects, physics is most influential
- Pass Irish is bad – no matter how well you do!
- CAO score has NO influence
- Neither does gender, living at home or extra subjects (i.e. not included in CAO score), or year of entry
- No difference between 2 engineering programmes
- Getting a lower CAO preference has an effect from about -0.02 to -0.2 (not shown on graph)
Evidence from the HEA study (2010) suggests that CAO points, English and Maths are best predictors.

We could make (a not very fair!) comparison, by constructing notional student profiles varying the above three variables from minimum to maximum (2x2x2=8) and within each of those categories differentiating between otherwise identical students with the extra information in this model.

- Thus – 8 pairs of students who would be identical by HEA predictors, but maximally different in our model.
### HEA study on progression

<table>
<thead>
<tr>
<th>CAO</th>
<th>Maths</th>
<th>English</th>
<th>Student</th>
<th>0.0%</th>
<th>20.0%</th>
<th>40.0%</th>
<th>60.0%</th>
<th>80.0%</th>
<th>100.0%</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>High</td>
<td>High</td>
<td>A</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>High</td>
<td>Low</td>
<td>B</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Low</td>
<td>High</td>
<td>C</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Low</td>
<td>Low</td>
<td>D</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>High</td>
<td>High</td>
<td>E</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>High</td>
<td>Low</td>
<td>F</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Low</td>
<td>High</td>
<td>G</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Low</td>
<td>Low</td>
<td>H</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>High</td>
<td>High</td>
<td>I</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>High</td>
<td>Low</td>
<td>J</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Low</td>
<td>High</td>
<td>K</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Low</td>
<td>Low</td>
<td>L</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>High</td>
<td>High</td>
<td>M</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>High</td>
<td>Low</td>
<td>N</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Low</td>
<td>High</td>
<td>O</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Low</td>
<td>Low</td>
<td>P</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

### How good are the predictions?

Let's take a quasi-philosophical example.

I predict that there is a 70% chance of rain tomorrow. If it rains (or not) – am I correct?

I can only ever be correct (or wrong) if I predict at the extremes, i.e. 100% or 0%.

I also have to decide to make a decision to use my prediction e.g. I might take an umbrella with me if the chances of rain are higher than 50%.

* The value of ‘X’ is going to depend on my relative tolerances for carrying a brolly on dry days versus getting wet when I don’t.

And a very good night t’ya!

The lovely Jean Byrne (a promised feather-clad ladies!)
Signal Detection Theory

A receiver operating characteristic (ROC) is a graphical way of displaying the compromise between false positives (which we need to avoid) and true positives (where there is a cost associated with false negatives)

At this point my sensitivity is 80%, but at a cost of a false positive rate of ~ 10%.
I'll be taking my brolly with me on 80% of rainy days, but on 10% of dry days I'll also have to lug it around!

No ability to discriminate — e.g.,
take my brolly with me X% (the climatological frequency of wet days) of the time

ROC - Engineering Progression

1. All Factors
2. Random
3. CAO
4. CAO + Math
What about the C3 requirement?

![Graph showing the relationship between CAO Multil Score and Probability of passing.]

Results – All of TCD

![Graph showing the size of effect for different courses.]

16
Results – All of TCD

Some observations

- Higher Maths has a positive effect, larger than any other subject
- Gender is significant
- So is faculty!
- Building construction and tech drawing have large effect – spatial awareness?
- Living at home is marginally beneficial
- Extra subjects (not included in best 6) contain very valuable information

Other uses of the data

Gender Uptake of Various Subjects

- Female
- Male
Other uses of the data

Discussion

General ATTRACT implications

- Poor understanding of engineering – especially amongst girls
- Need to breach the gap between ‘it is good’ and ‘but not for me’
- Significant biases exist within school system that steer girls away from STEM (biology the exception)
- Teachers and parents need to be a focus of our outreach
- Active learning/participative activities are most successful – consider how we could use the web here maybe?
- First 6 weeks are critical in retention
- Need to address both academic and social integration
Discussion

Data Analysis implications

- There is considerable information in the LC grades that could be applied at entry time
  - Assume we want to match ability to progress with entrance?
  - Need to balance transparency and ease of understanding
- We can also use this information post entry – identify those most likely to be at risk of failing to progress and target supportive action
  - Consider that a first year student ‘saved’ brings funding to college at low/minimal marginal extra cost
  - This benefit continues through all subsequent years
  - Can we afford not to do this?

A simple economic model

<table>
<thead>
<tr>
<th>Metric</th>
<th>Value</th>
<th>Notes</th>
</tr>
</thead>
<tbody>
<tr>
<td>Dropout Rate in 2nd year</td>
<td>20%</td>
<td>Percentage of students dropping out in the 2nd year without academic records</td>
</tr>
<tr>
<td>Number of Students</td>
<td>100</td>
<td>Total number of students enrolled in the program</td>
</tr>
<tr>
<td>Ex-Post Student Exp</td>
<td>€11,000</td>
<td>Total cost to students after the elimination of financial support</td>
</tr>
<tr>
<td>Cost per student revealed to college</td>
<td>€19,000</td>
<td>Total cost to college after the elimination of financial support</td>
</tr>
<tr>
<td>% of students targeted</td>
<td>20</td>
<td>Percentage of students who are targeted by internal initiatives</td>
</tr>
</tbody>
</table>

| Calculating students | 50 | Number of students being calculated |
| Students | 50 | Total number of students being targeted |
| Students to target | 30 | Percentage of students targeted by initiative |
| Students targeted | 25,000 | Percentage of targeted students who are actually targeted |

Economic Benefit of Targeted Retention Initiative

- % decreases left to determine profit
- Various negative outcomes ranging from €10,000 to €50,000
- % increases left to determine profit
- Various positive outcomes ranging from €50,000 to €100,000

Percentage of Total Students Targeted by Initiative

| ATTRACTION | Context | Entrance | Data | Seeing the Future - 24th January 2013 | Analysis | Results | Discussion | Future |
Future Work

Publishing
- A lot done, a lot more to do!

Analysis
- Extension of LC analysis to all LC students (data for 10-11 and 11-12 obtained from CAO)
- Look at project maths influence

Outreach
- Review and revise outreach activities

More research
- Look at opportunities in Horizon 2020

Action
- Shouldn’t we in TCD be doing something with this knowledge?

Thank you!