Lies, damned lies, SET & marks

Using statistical analysis of student marks & feedback to characterise module assessment performance

Tom Reader
School of Life Sciences
University of Nottingham
Outline

• Background: HE goes metric
• Student evaluation: what does it tell us?
• Marks data: three ways to describe a module
• Diagnosing and fixing marks problems
• Do marks predict student evaluations?
HE goes metric

• Student evaluation & NSS determine University strategy, funding & performance review
• Marks data at the heart of what we do
• TEF to include metrics of satisfaction & value added?
• Marks data used to measure value added?
What does student evaluation tell us?

• Very widely used (reviewed in Spooren et al 2013)
• One of very few measures of teaching “quality”
• Hard to interpret
• Beauty contest?
• Grade inflation?
• Negative relationship with teaching effectiveness (Braga et al 2014)
• NSS does not predict medical degree outcomes (Lancaster & Fanshawe 2015)
What do module marks tell us?

Signal
- Student ability
- Student effort

Noise
- Subject difficulty/interest
- Assessment type
- Inter-assessor variance
- Intra-assessor variance
- Mark scheme
- Clerical errors

Legitimate?
Understandable?
Problematic?

Bridges et al (2002)
Tian (2005)
Research aims & methods

• Systematic analysis of student evaluation & module marks data
• Identify shortcomings of data
• Find positive use for data in diagnosing issues with modules

• Data from School of Life Sciences, University of Nottingham
• Principal components analysis, correlation & GLMs
• Visual Basic, R & SPSS
Student evaluation @ Nottingham

SEM:
1) The objectives of the module were made clear/available to me
2) The teaching methods used in the module help me to learn
3) The method of assessment allows students to demonstrate what they have learned
4) The level of the module is about right
5) The workload on the module is about right

SET:
1) The teacher was an able communicator
2) The teacher retained my interest
3) The teacher was approachable
4) Sessions were paced appropriately
5) Overall, this teacher assisted my learning

Answers:
5 – strongly agree
4 – agree
3 – neutral
2 – disagree
1 – strongly disagree
SET and SEM are related

\[ p < 0.001 \]
\[ r = 0.529 \]

Aggregate scores
Student evaluation – analysis

Do they tell us 10 different things?

Principal components analysis:

<table>
<thead>
<tr>
<th></th>
<th>Eigen value</th>
<th>% of Variance</th>
<th>Cumulative %</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>5.422</td>
<td>54.2</td>
<td>54.2</td>
</tr>
<tr>
<td>2</td>
<td>2.009</td>
<td>20.1</td>
<td>74.3</td>
</tr>
<tr>
<td>3</td>
<td>0.795</td>
<td>7.9</td>
<td>82.3</td>
</tr>
<tr>
<td>4</td>
<td>0.502</td>
<td>5.0</td>
<td>87.3</td>
</tr>
<tr>
<td>5</td>
<td>0.415</td>
<td>4.2</td>
<td>91.4</td>
</tr>
<tr>
<td>6</td>
<td>0.360</td>
<td>3.6</td>
<td>95.0</td>
</tr>
<tr>
<td>7</td>
<td>0.186</td>
<td>1.9</td>
<td>96.9</td>
</tr>
<tr>
<td>8</td>
<td>0.136</td>
<td>1.4</td>
<td>98.2</td>
</tr>
<tr>
<td>9</td>
<td>0.096</td>
<td>1.0</td>
<td>99.2</td>
</tr>
<tr>
<td>10</td>
<td>0.080</td>
<td>0.8</td>
<td>100.0</td>
</tr>
</tbody>
</table>

Sample:
- 89 modules
- 6601 students
Student evaluation – analysis

Student evaluation tells us two things:

<table>
<thead>
<tr>
<th></th>
<th>PC1</th>
<th>PC2</th>
</tr>
</thead>
<tbody>
<tr>
<td>q1</td>
<td>0.728</td>
<td>0.356</td>
</tr>
<tr>
<td>q2</td>
<td>0.816</td>
<td>0.388</td>
</tr>
<tr>
<td>q3</td>
<td>0.558</td>
<td>0.507</td>
</tr>
<tr>
<td>q4</td>
<td>0.672</td>
<td>0.613</td>
</tr>
<tr>
<td>q5</td>
<td>0.435</td>
<td>0.593</td>
</tr>
</tbody>
</table>

PC1:  + I am happy
      - I am unhappy

PC2:  + I am happier with module than lecturer
      - I am less happy with module than lecturer
### Student evaluation – analysis

**Basic patterns:**

Mean completion rate: 35% (lower than paper-based)

**Associations (General Linear Model):**

<table>
<thead>
<tr>
<th></th>
<th>PC1</th>
<th>PC2</th>
</tr>
</thead>
<tbody>
<tr>
<td>Year</td>
<td>p &lt; 0.05</td>
<td>NS</td>
</tr>
<tr>
<td>Completion rate</td>
<td>NS</td>
<td>NS</td>
</tr>
<tr>
<td>Class size</td>
<td>NS</td>
<td>NS</td>
</tr>
<tr>
<td>Teaching group</td>
<td>NS</td>
<td>NS</td>
</tr>
</tbody>
</table>

**Sample:**
- 89 modules
- 6601 students
Student evaluation – analysis

![Graph showing mean PC1 over years](image-url)
Marks data – acquisition

- Raw marks of limited value
- Do high marks reflect students or module?
- Better: Student performance relative to other modules

<table>
<thead>
<tr>
<th>Student</th>
<th>Mark on module X</th>
<th>Mean on other modules</th>
<th>Deviation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Fred</td>
<td>50</td>
<td>55</td>
<td>+5 %</td>
</tr>
<tr>
<td>Sarah</td>
<td>67</td>
<td>73</td>
<td>-6 %</td>
</tr>
<tr>
<td>Rachael</td>
<td>72</td>
<td>62</td>
<td>+10 %</td>
</tr>
<tr>
<td>Abdi</td>
<td>70</td>
<td>68</td>
<td>+2 %</td>
</tr>
<tr>
<td>Emma</td>
<td>53</td>
<td>50</td>
<td>+3 %</td>
</tr>
</tbody>
</table>

Mean: +2.8 %
How typical is a module?

Mean deviation = relative performance
How typical is a module?

Module mark

Mean mark on all other modules

$R^2 = \text{predictability}$

Slope = discrimination
Relative performance

Sample:  
- 56 modules
Relative performance

Mean deviation (%) +/- SEM

Tutorial modules
Relative performance

Practical skills

Mean deviation (%) +/ - SEM
Relative performance

Final year project

Mean deviation (%) +/− SEM
Relative performance

“Hard” modules?
### Predictability

#### Bell weather

#### Random?
Discrimination

Discriminatory?

Non-discriminatory?

Mean mark excluding module

Mark on module
How typical is a module?

Non-discriminatory?

“Easy” modules?

“Hard” modules?
Fixing problems

Removing a clerical error
Fixing problems

Dealing with a problematic exam
An unfair advantage?

The advantage of cherry picking modules is small.
Exams versus coursework

Exams correlate with discrimination

$R = 0.598; \ p < 0.001$
Exams versus coursework

Exams correlate with predictability

$R = 0.502; \ p < 0.001$
Exams versus coursework

Exams correlate with relative performance

R = 0.503; p < 0.001
Does evaluation reflect marks?

- Principal Components Analysis on SEM data
- Pearson’s correlations

- Note: SEM performed **before** exams but **after** some coursework marks
Does evaluation reflect marks?

No relationship between evaluation & discrimination

$r = -0.212; p > 0.05$
Does evaluation reflect marks?

No relationship between evaluation & predictability

$r = -0.214; p > 0.05$
Does evaluation reflect marks?

Students like the modules in which they do well

$r = 0.54; p < 0.001$
Does evaluation reflect marks?

...but more so when there is coursework

ANOVA: Interaction: $F_{(1,38)} = 5.508; p < 0.05$
Does evaluation reflect marks?

Coursework modules not more popular

\[ r = -0.082; \ p > 0.05 \]
Does evaluation reflect marks?

Practical modules more popular? (n = 3)

\[ F_{(2,39)} = 1.914; \ p = 0.161 \]
Conclusions

SET & SEM overlap strongly
Student evaluations are not independent of marks
Collation & analysis of marks data is:
• Feasible
• Useful
• Essential?

But...
• Modules vary for legitimate reasons
• Student achievement is multivariate
• Quality of data available in other schools/institutions?
Acknowledgements

- Our external examiners
- My colleagues
- School of Life Sciences Undergraduate Office

References