

How was it for you?
Psychophysics and the evaluation of student experience of e-learning

Diana Eugenie Kornbrot, University of Hertfordshire
d.e.kornbrot@herts.ac.uk

Evaluating the student experience of Higher Education has become a matter of national importance in several countries. For example, in England & Wales the National Student Survey (NSS) is administered on line to all students in the final year of their undergraduate degree. The NSS uses 5-point Likert scales, giving extent of agreement or disagreement with positive statements. There are 22 questions, covering 6 aspects of student experience. This presentation considers how psychophysical methods based on signal detection theory or Luce's choice theory can be used to analyze such data. Such methods can determine how well the questions discriminate different aspects of experience, as well as how favourably the students experience these aspects of their education. Particular emphasis is given to exploring discipline differences together with the effects of recent technologies, such as managed learning environments and web 2.0 social software.

Higher education matters: to students, parents, communities, countries, the planet. Consequently, everyone wants to evaluate the success of the higher education enterprise. This Ms. draws on both performance and satisfaction data from higher education (university) data on satisfaction and performance in England & Wales to illustrate the problem and show how psychophysics can make a contribution. There are two key reasons why evaluation remains so difficult. Namely, such evaluations are 'high stake' and have multiple stakeholders.

High stakes are manifest in the importance of a good degree. For the individual student, getting a 'good' degree is of paramount importance, preferably with as little effort and risk as possible. This poses a fundamental problem for people devising assessment procedures. Identical procedures do *not* measure the same underlying property over time. IQ tests provide an interesting example, with massive improvements over the the C20th (Flynn, 1987). As achieving a high IQ is 'high stake', 'teaching to the test' became rife. Many cried 'foul', and tried: either to restrict such teaching, or to devise non-coachable items. Doomed to failure. If the results are high stake, then there is an inevitable 'arms race' between the assessors and the assesseees. What the psychological instruments measure changes. Physicists don't have this problem. The ruler measures length, the scales measure weight, etc..

The problem is just as intractable when assessing student satisfaction. All the stakeholders would like to know how students experience features of education such as: teaching, assessment, learning resources, personal development, etc.. Clearly here too, if it is in a student's interest to graduate from a high satisfaction institution, then student may modify their questionnaire responses in line with those interests.

Higher education evaluation has multiple stakeholders. Degree results are widely used to validate psychological properties that predict how people will perform on *future* tasks. This is equally true of the potential graduates who will be the actual task performers and of the employers who will profit from their efforts. It is also true of the communities, local, national and international that will benefit from a high quality graduate workforce. Furthermore, the paying clients of the higher education industry, be they students, parents, government or industry will want 'value for money'. In attempting to satisfy these needs higher education evaluation is actually fulfilling two, sometimes conflicting, purposes. Firstly, degree performance informs the world about what the student *already* knows. Doctors know about the functioning of the human body, physicists about Newton's Laws and

quantum electrodynamics, computer scientists can program in at least one language, etc. Any job one can think of will require *some* existing knowledge. Secondly, degree performance informs the world about the graduates ability to learn *new* knowledge quickly. Attempting to measure both psychological properties at the same time is fraught with problems. However, that's what the stakeholders want or need. Indeed employers depend heavily on universities for their selection processes. Thus degree performance measures need to recognize the following stakeholder needs and conflicts. Firstly students have a stake in obtaining a 'good' degree, but only if that degree is recognized as being tough. A good degree is useless if obtained by every student. Secondly, universities have a stake in increasing the proportion of students with good degrees – up to a point. Thirdly, employers have a stake in knowing what a 'good' degree means. Similar arguments apply to student satisfaction. The stakeholders are the same as with degree performance. Employers want graduates from institutions where students were satisfied with teaching and assessment.

In my view the most important substantive issue is the determination of how performance and satisfaction depend on discipline (subject). Both performance and satisfaction will depend on the discipline (subject) studied. If engineers are less satisfied than physicists, as in the UK, then universities with lots of physicists & few engineers will have better satisfaction scores. If philosophers achieve a smaller proportion of excellent degrees than physicists, as in the UK, then universities with lots of physicists & few philosophers will have better degree performance scores. Only once discipline base lines have been established, can one go on to determine how satisfaction and performance depend on other key factors: change over the years or decades, incoming achievement of the students and identity or type of higher education institution.

There are also important methodological issues, where psychophysics has the potential to contribute. In the spirit of psychophysics, one can compare discipline performance in terms of both bias and discriminability. i.e. Is the difference between physics and philosophy merely one of criterion shift, or are the criteria spacing different? Similarly, are within discipline differences across time simply criteria shifts; or are there different relative criteria spacings.. In terms of satisfaction, one can compare different features (teaching, assessment, resources, etc) both within and across disciplines.

Degree performance and incoming grades of students for England & Wales have been held centrally for many years. More recently, a National Student Satisfaction (NSS) questionnaire has been administered to all final year undergraduates. These data sets are used to address. This data has been used to address both substantive and methodological issues.

Data Analysis

In the UK most undergraduates take an honours first degree in one major subject. Degree performance is then graded as one of the following: 1st, 2.1, 2.2, 3rd, pass or fail. The Higher Education Statistics Association, HESA, has proportion of students achieving all classifications for each Higher Education Institution, HEI, separately for each of some 160 + subject codes (HESA, 2007). From these data the measures of performance were constructed for 18 narrow discipline groups. The measures are defined as the % achieving the following criteria: 1st class defined as 'excellent' 'good', 1st or 2.1, defined as 'good'; 1st or 2.1 or 2.2 defined as 'competent'. A good degree is typically required for postgraduate research work or employment in prestigious companies. Thus a good degree is a career passport for a student or a kite mark for an employer.

It was possible to identify 18 narrow disciplines, in four broad areas. Just 12 are used here (for simplicity). They are: **physical sciences**, *high maths* (imaths, physics, chemistry),

low maths (others such as geology), *engineering* (including computer science); **biological sciences**, *biolog* (including molecular, microbiology, anatomy), *human science*, (including psychology, linguistics, anthropology), *health* (professions allied to medicine, etc), *caring* (nursing, teaching, social work); **social sciences**, *quantitative* (economics, finance, etc), *qualitative* (sociology, politics, etc), *business*; **humanities & arts**, *arts* (languages, literature, history, philosophy, etc), *verbal communication* (journalism, etc), visual arts (graphics, painting, design, etc), *performance* (drama, music). These classifications were made starting from the traditional definitions and then grouping together disciplines with very similar satisfaction patterns. This procedure put psychology with linguistics, not biology and nursing with teaching rather than other applied biology. This data for 06-07 was analyzed according to discipline. Data for high maths, engineering, biology and arts were also compared with similar data from 1980-82 (Kornbrot, 1987).

Data on the performance of incoming students was obtained from the University Central Admissions Service (UCAS). They use a points system where 480 points corresponds to 4 A grades at High School graduation level. The measure used here is mean points/480 times 100 to be comparable with the degree results. A score near 80% corresponds to at least AAB and is excellent (e.g. more maths incoming students). A score near 70% corresponds to at least 2As or BBC and is also high performance (e.g. other academic disciplines). A minimum of EE equivalent to 17% is required for university entrance. 40%, the lowest average in our data corresponds to DDE.

Student satisfaction was measured using the National Student Survey. This asks students whether they strongly agree, agree, neither agree nor disagree, disagree or strongly disagree with the 22, classified into 7 themes: teaching, assessment & feedback, academic support, organization & management, learning resources, personal development and overall experience see (Richardson, Slater, & Wilson, 2007) for development of the instrument. Data for 2006-7 were obtained from the National Student Survey web site <http://www.thestudentsurvey.com/>.

Results

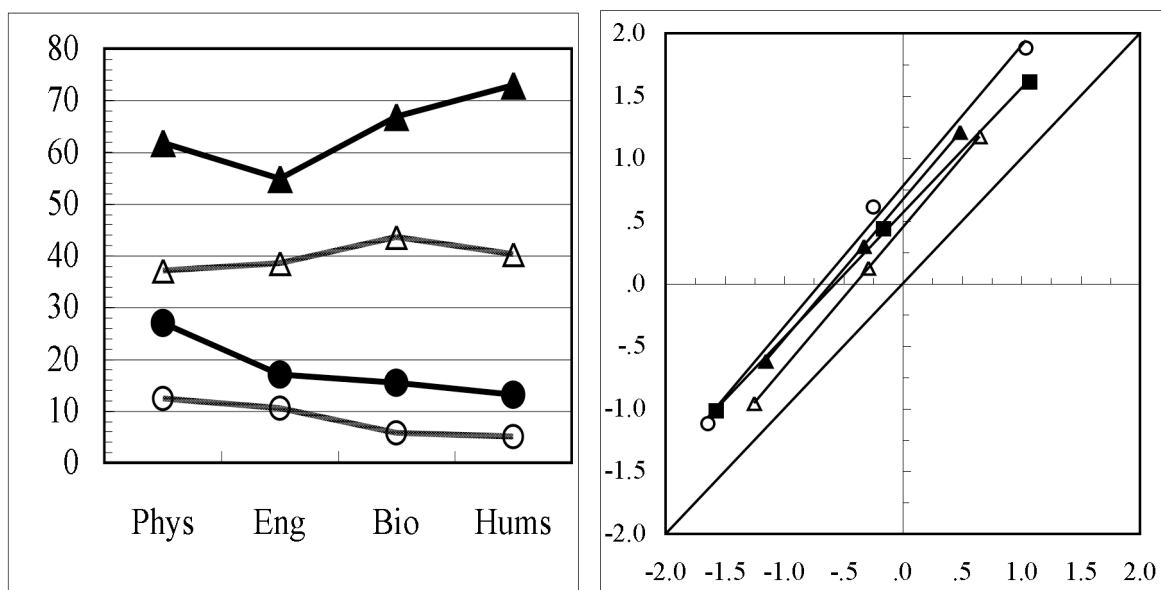


Figure 1. Inflation. Left panel shows percent achieving degree type as by discipline: circles = excellent; triangles = good. Open symbols 80-82, filled symbols 06-07. Right panel shows z-scores in 06-07 as function of z-scores in 80-82: filled triangles=physics, filled squares = biology; open triangles = engineering; open triangles =humanities.

Figure 1 shows the effect of time on degree performance. The left panel shows percent achieving good and excellent degrees for 06-07 and 80-82. The increase in percentage terms is substantial, and appears larger for good than excellent performance. The right panel takes a Thurstonian approach based on cumulative probabilities. It shows z-scores for excellent, good and competent achievement in 06-07 as a function of equivalent values in 80-82. The shift of all lines above the diagonal (no change line) show the improvement of performance over 25 years: physics = .68, biological sciences = .56, engineering = .45, humanities = .78. The functions are all highly linear, with adjusted r^2 for humanities .992, and all other adjusted r^2 greater than .999. Engineering and physics had slopes slightly above 1, indicating the increase in performance was slightly lower for the more demanding excellent criterion.

Fig. 2 shows the effect of discipline, with narrow discipline groups, grouped into the four broader areas, together with incoming grades. In each broad area, the more academic disciplines have higher proportions of all levels of achievement, not surprising as the incoming students have better pre-university performance. In the UK system this generally means these disciplines are more popular, as students have to achieve higher grades to be accepted. Nevertheless, higher incoming grades does not necessarily lead to higher performance. In physical science, biological and social science higher incoming grades leads to more excellent performance, but no difference in good performance. This is true when comparing *more maths* with *less maths*; when comparing *biology* with *human sciences*; and when comparing *quantitative* with more *qualitative*. Conversely, applied disciplines have higher percentage of excellent than less popular academic disciplines, even though their incoming grades are *lower*. This is true when comparing *engineering* with *low maths*; *health & caring* with *human sciences*; and *business* with *qualitative*. In all these cases the percentage of good degrees follows incoming grades, and is higher for the academic discipline. These findings are supported by z-score analyses. Regressing the highest performing discipline on other disciplines in the same area gives highly linear functions, with minimum adjusted r^2 of .994 and most greater than .999. All regressions have positive intercept corresponding to higher percentages achieving all grades. If slopes are different from unity they are lower, indicating less difference at excellent than competent level.

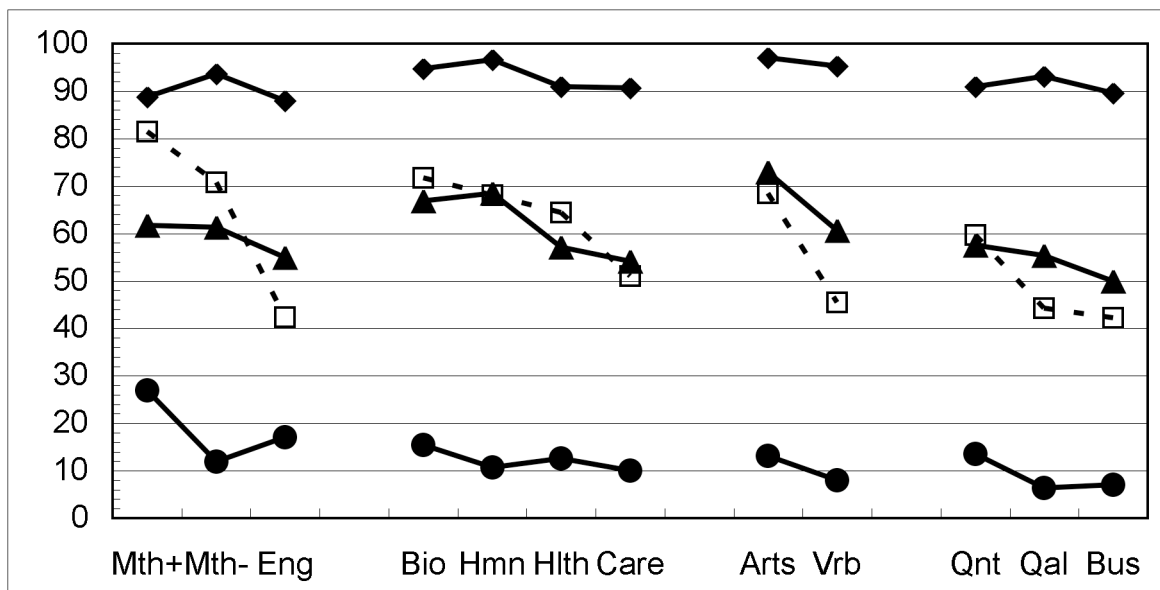


Figure 2. Discipline effects. Percentage achieving degree category as a function of discipline: filled circles= excellent; filled triangles=good. Open circles show mean incoming student performance as a percentage of points for 4 Agrades.

Figure 3 shows satisfaction as a function of discipline.

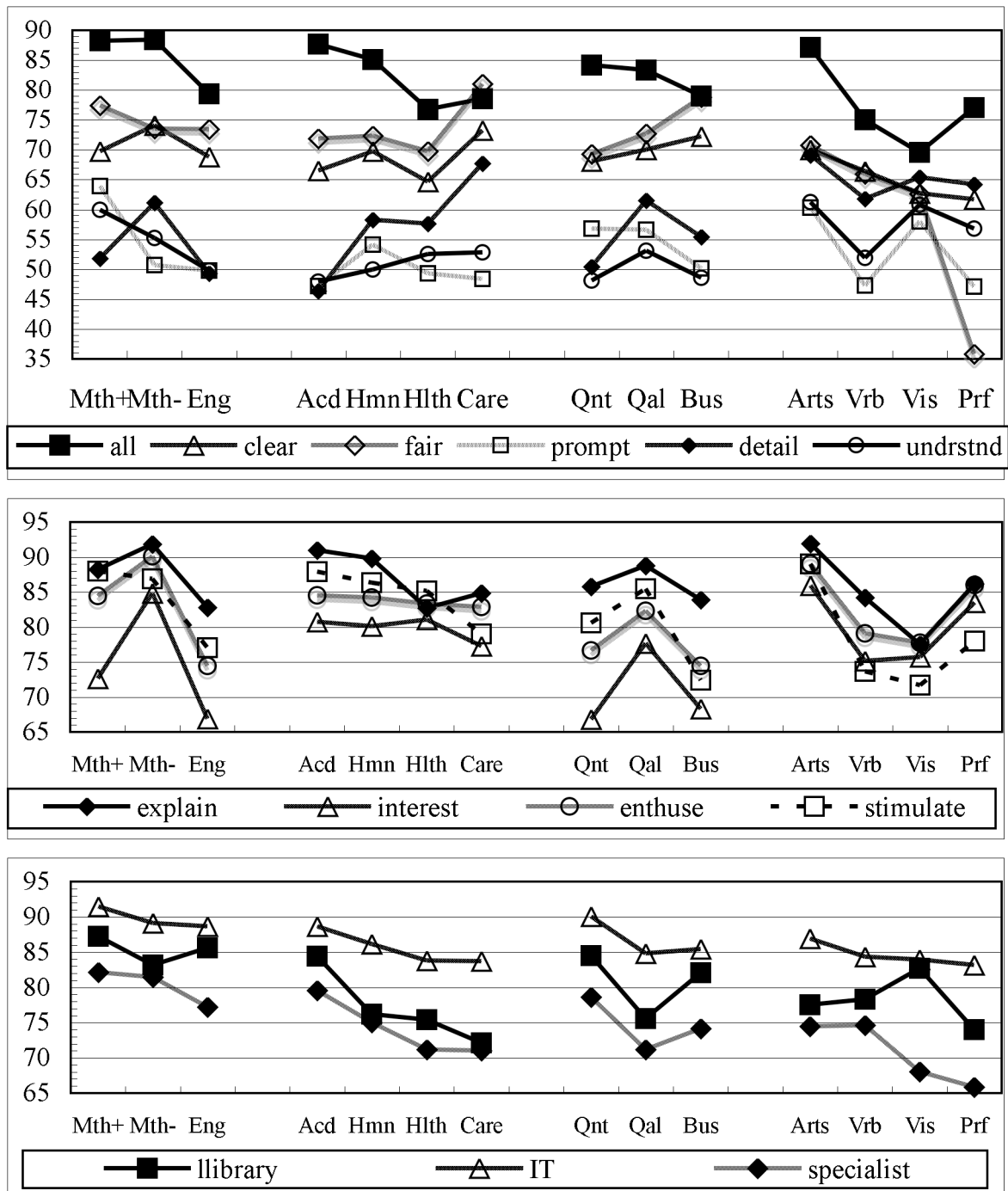


Figure 3. Student satisfaction. Percent agreeing with positive statement as a function of discipline. Panels: top overall and assessment; middle, teaching; bottom learning resources.

The top panel shows a high level of overall satisfaction, but it is striking that all academic disciplines are above, and all applied disciplines are below 80%. Assessment is viewed less favourably, particularly in its contribution to understanding and promptness. Assessment detail is also relatively low, and even where it is relatively high (*health, care*), it is not associated with understanding. The middle panel shows teaching explanations are generally high, except in, *high maths* and *quantitative*. Applied disciplines, with lower incoming grades, rate explanation less highly. Most disciplines are more stimulating than interesting,

with a general pattern following that of explanation. Lecturer enthusiasm was lower in applied disciplines. The bottom panel shows across the board high satisfaction with IT resources.

Discussion

Clearly, there have been massive change in the proportion of students achieving excellent and good degrees over the past quarter of a century. It seems very unlikely that this is due to a massive change in student intelligence. Conversely, independent examiner evaluation of difficulty of questions and quality of answers does not show any concomitant change. One can only conclude that similar instruments are measuring different properties. This presents a quandary, employers would probably like to identify the top 40%, as indicated by a good degree in 1980, but cannot insist on a 1st, as there aren't enough available. They are left relying on personal academic references to distinguish among 2.1 students, obviously far less reliable than the externally validated degree awarding process. This is also a major problem for graduates, since their achievements are undermined. It also militates against students with good degrees from less prestigious universities, often working class, as employers no longer take a student with a good degree from lesser universities, since such a high proportion from more prestigious universities have good degrees.

The satisfaction scores, in spite of their shortcomings, have a dramatic message about what universities are doing well, and what they are doing less well. Features that depend on a high staff student ratio, particularly assessment that is detailed, timely and contributes to student understanding got disappointingly low ratings. Features that can be delivered almost equally well to 300 or 30 students, particularly teaching explanations were highly rated. IT was also highly related, presumably the drop in PC prices has offset the increase in student numbers.

The discipline differences are also striking. In this supposedly materialistic age, the strong preference for more academic disciplines was perhaps unexpected. The superior performance and satisfaction in these disciplines may be due to students studying their preferred subject, as well as to their superior incoming achievement. Science, particularly physical science maintains a higher proportion of excellent degrees with a relatively lower proportion of good degrees, as compared with humanities and to a lesser extent social sciences.

The Thurstonian analyses of performance shows interesting similarities across time and disciplines. If criteria had stayed constant, the observed shifts would be equivalent to an increase in IQ of 7 to 12 points. Since this is unlikely, it is probable that criteria have shifted. Interpreting such results and devising insightful evaluations remains a key research problem.

References

- Flynn, J. R. (1987). Massive IQ gains in 14 nations: What IQ tests really measure (Vol. 101, pp. 171-191).
- HESA. (2007). Higher Education Statistics Agency: Statistics publications [Electronic Version] from <http://www.hesa.ac.uk/products/pubs/home.htm>.
- Kornbrot, D. E. (1987). Degree performance as a function of discipline studied, parental occupation and gender. *Higher Education*, 16, 513-534.
- Richardson, J. T. E., Slater, J. B., & Wilson, J. (2007). The National Student Survey: development, findings and implications. *Studies in Higher Education*, 32(5), 557 - 580. <http://www.informaworld.com/10.1080/03075070701573757>