



CAMBRIDGE ASSESSMENT

# **The predictive validity of the Thinking Skills Assessment: a combined analysis of three cohorts**

*Joanne L Emery*

**03 October 2007**



## Background and Summary

This report looks at the predictive validity of the Thinking Skills Assessment (TSA) for 1<sup>st</sup> year examination performance in Computer Science, Economics, Engineering and Natural Sciences at the University of Cambridge. Three successive years of TSA and examination data were combined using standardized scores and analysed as one. Correlations, scattergrams and logistic regression plots are presented in support of the test. Correlation coefficients and scattergrams suggest a moderate positive relationship between TSA scores and examination performance. Logistic regression plots suggest that total TSA scores clearly predict the probabilities of achieving the highest and the lowest examination classes.

## Methods and Statistics

TSA data from 2003, 2004 and 2005 were merged with subsequent 1<sup>st</sup> year examination data supplied by the University of Cambridge. Data were matched by name and then anonymised. TSA performance is reported as a total calibrated score and as separate Problem Solving and Critical Thinking sub-scores. The calibration process allows the results of candidates taking different test versions to be reported on a common interval scale. Problem Solving questions assess reasoning using numerical and spatial skills. Critical Thinking questions assess reasoning using everyday written language. All questions are multiple-choice and marking is therefore objective. For details of the test, please refer to the website: [www.tsa.ucles.org.uk](http://www.tsa.ucles.org.uk)

Total marks for 1st year examinations from 2005, 2006 and 2007 were obtained for Computer Science, Economics, Engineering and Natural Sciences. Marks were standardized (normalized) for each subject and each cohort using z-scores prior to data matching. This process converts raw scores into standard deviation units above or below the mean and allows the combination of observations from different cohorts. A positive z-score applies to raw scores above the mean and a negative z-score applies to those below it. After matching with the TSA data, the three cohorts were combined and analysed as one. The rationale for combining the cohorts is that predictive validity coefficients often vary widely between successive cohorts at the same institution. Table 1 shows the number of matches made for each course and cohort.

Table 1: Numbers of TSA candidates matched to 1<sup>st</sup> year examination data in each cohort

course	TSA cohort	exams taken	n matches
Computer Science	2003	2005	67
	2004	2006	47
	2005	2007	56
Economics	2003	2005	25
	2004	2006	54
	2005	2007	65
Engineering	2003	2005	105
	2004	2006	121
	2005	2007	143
Natural Sciences	2003	2005	106
	2004	2006	226
	2005	2007	245

Pearson correlations were calculated between TSA scores and examination total mark z-scores for Computer Science, Engineering and Natural Sciences. Distributions of z-scores for Economics were skewed and Spearman correlations were therefore more appropriate in this case. Scattergrams were plotted to accompany these. Hypotheses were one-tailed as a positive relationship was expected.

Accepted TSA candidates in each cohort have a higher mean and a narrower range of test scores than the cohort as a whole (see previous reports available on the TSA website). This is a problem for correlational analysis as it tends to restrict the magnitude of the coefficients. There are corrective formulae for range-restriction (Sackett and Yang, 2000) but these are only appropriate in particular circumstances and should not be routinely applied when the selection process is complex. Thus simple, uncorrected coefficients are presented throughout this report and these can be interpreted with reference to Table 2.

Table 2: Guidelines for interpreting uncorrected correlation coefficients in predictive validity studies

Validity Coefficient	Interpretation
Above 0.35	very beneficial
0.21 to 0.35	likely to be useful
0.11 to 0.20	depends on circumstances
Below 0.11	unlikely to be useful

Source: US Department of Labor, *Employment Training and Administration*, 1999

Logistic regression may be a more suitable method for assessing predictive validity as it does not make the data assumptions of correlation (linearity and normality). This technique predicts the probability of a discrete/binary outcome (here examination class) as a function of predictor variables. Students are awarded an examination class based on their overall total mark. Classes are awarded normatively each year rather than being based on pre-set cut-scores (a certain proportion of candidates are awarded each class). Classes, in order of descending merit, are as follows: Class 1, Class 2 (division 1), Class 2 (division 2), Class 3, Ordinary Pass, Fail. The Natural Sciences course does not subdivide Class 2 outcomes. Logistic regression analysis was carried out on the probability of achieving a 1<sup>st</sup> class outcome and on the probability of achieving a 3<sup>rd</sup> class outcome or below (including fails and withdrawals).

## Results

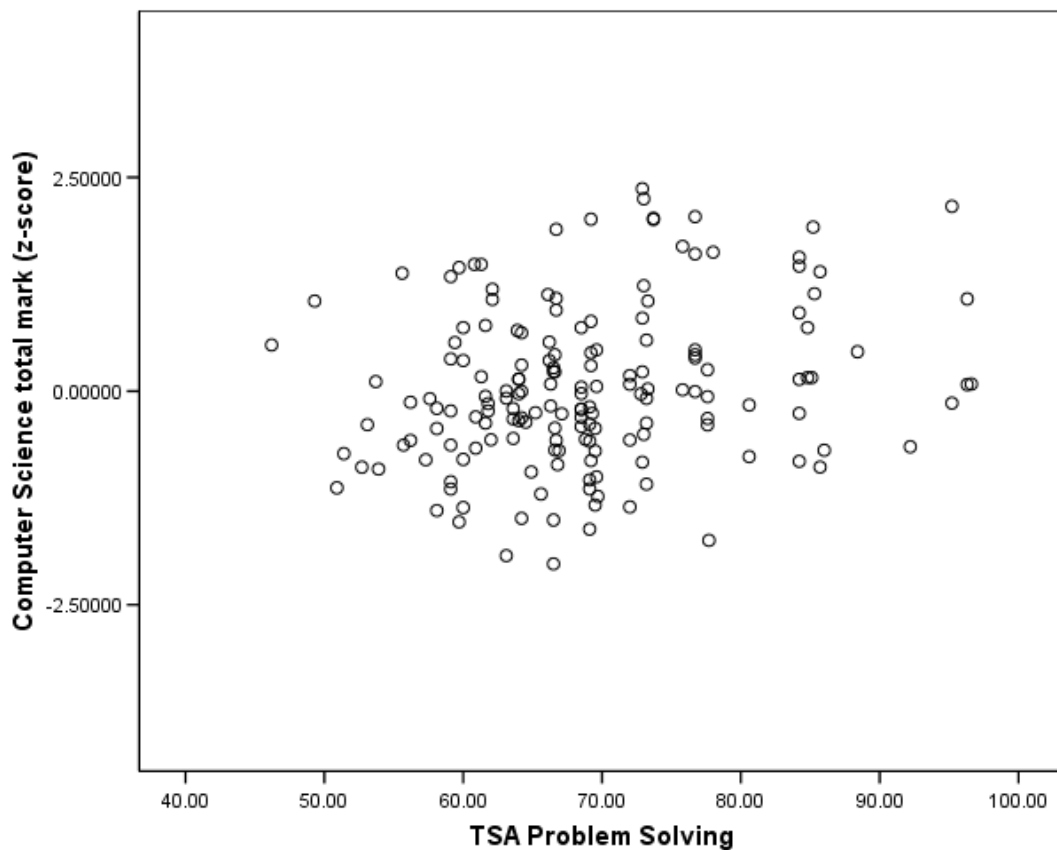
Table 2: Correlation coefficients between TSA scores and 1<sup>st</sup> year examination performance.

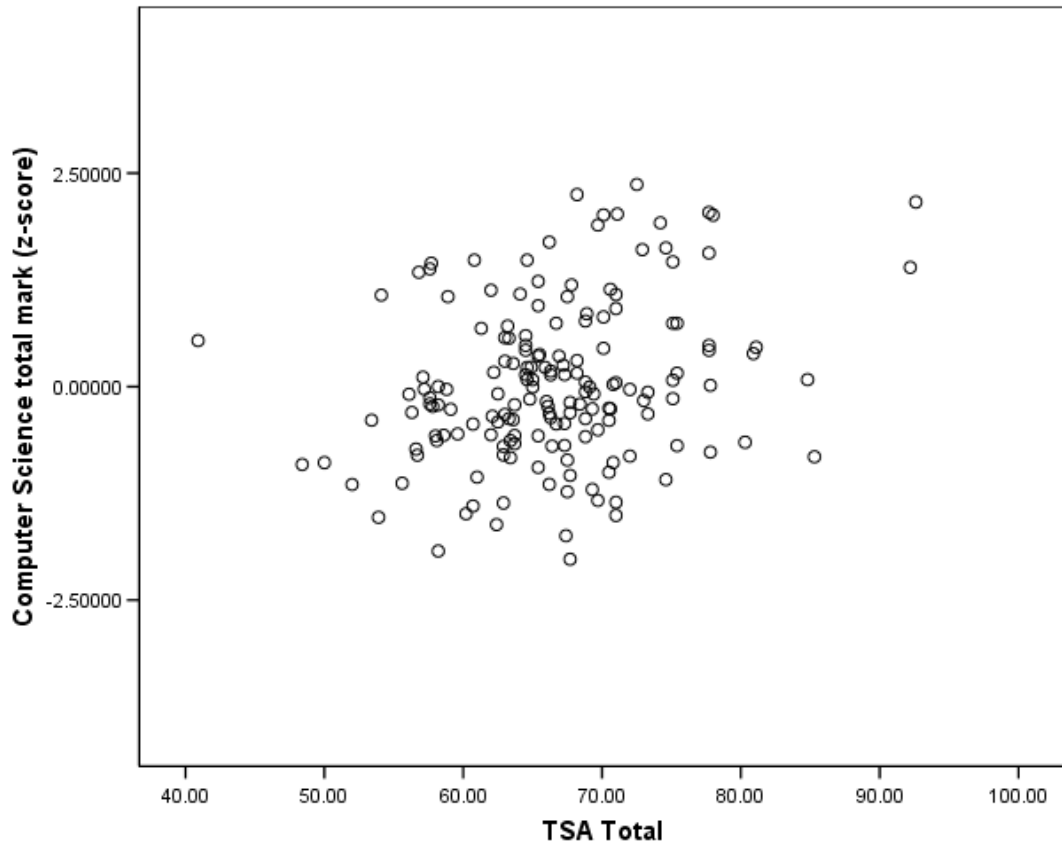
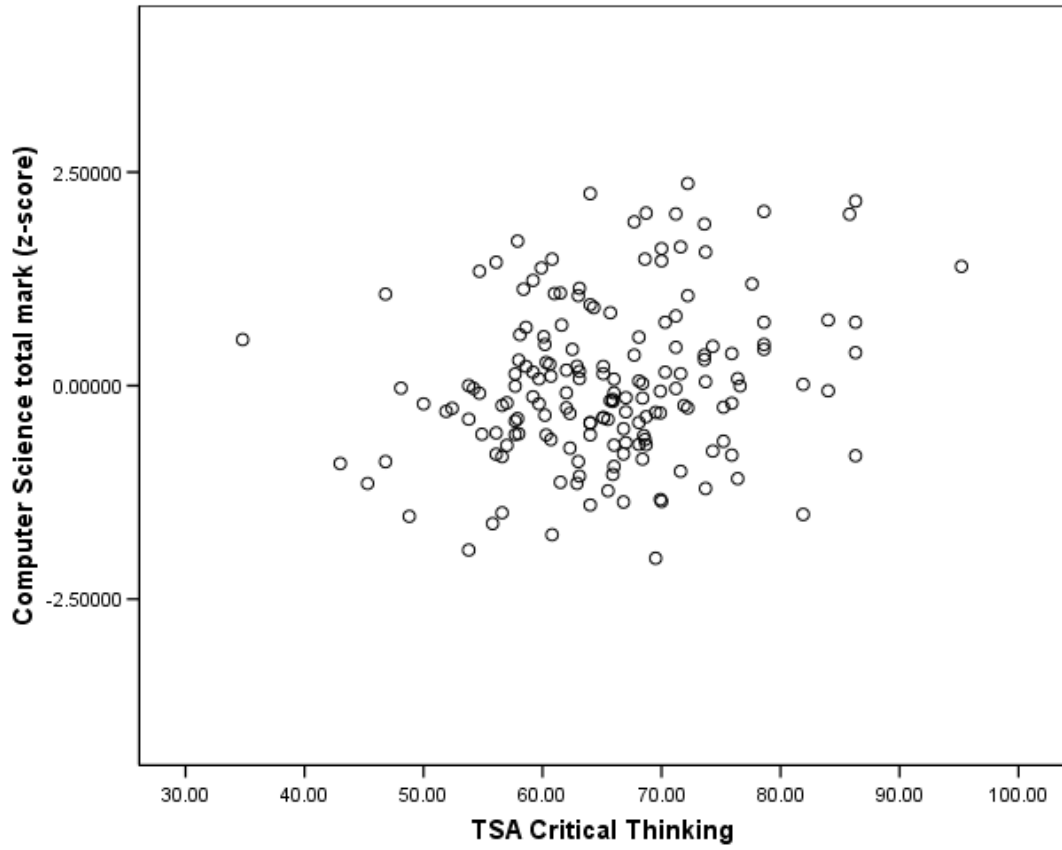
		TSA Problem Solving Score	TSA Critical Thinking Score	TSA Total Score
Computer Science total mark (z-score) n=167	Pearson coefficient	0.227**	0.216**	0.268**
	significance (one-tailed)	0.002	0.003	<0.001
Economics total mark (z-score) n=141	Spearman coefficient	0.275**	0.205**	0.281**
	significance (one-tailed)	<0.001	0.007	<0.001
Engineering total mark (z-score) n=368	Pearson coefficient	0.278**	0.182**	0.268**
	significance (one-tailed)	<0.001	<0.001	<0.001
Natural Sciences total mark (z-score) n=575	Pearson coefficient	0.193**	0.128**	0.189**
	significance (one-tailed)	<0.001	0.001	<0.001

\*\* significant at the 0.01 level

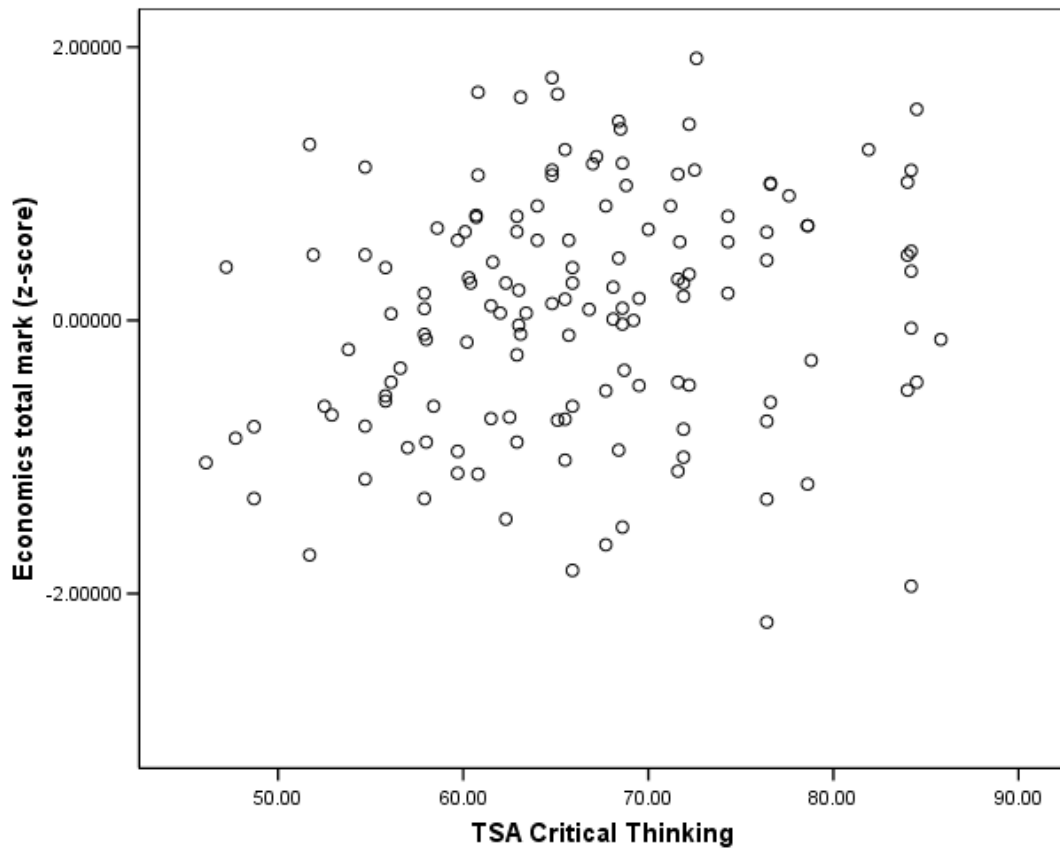
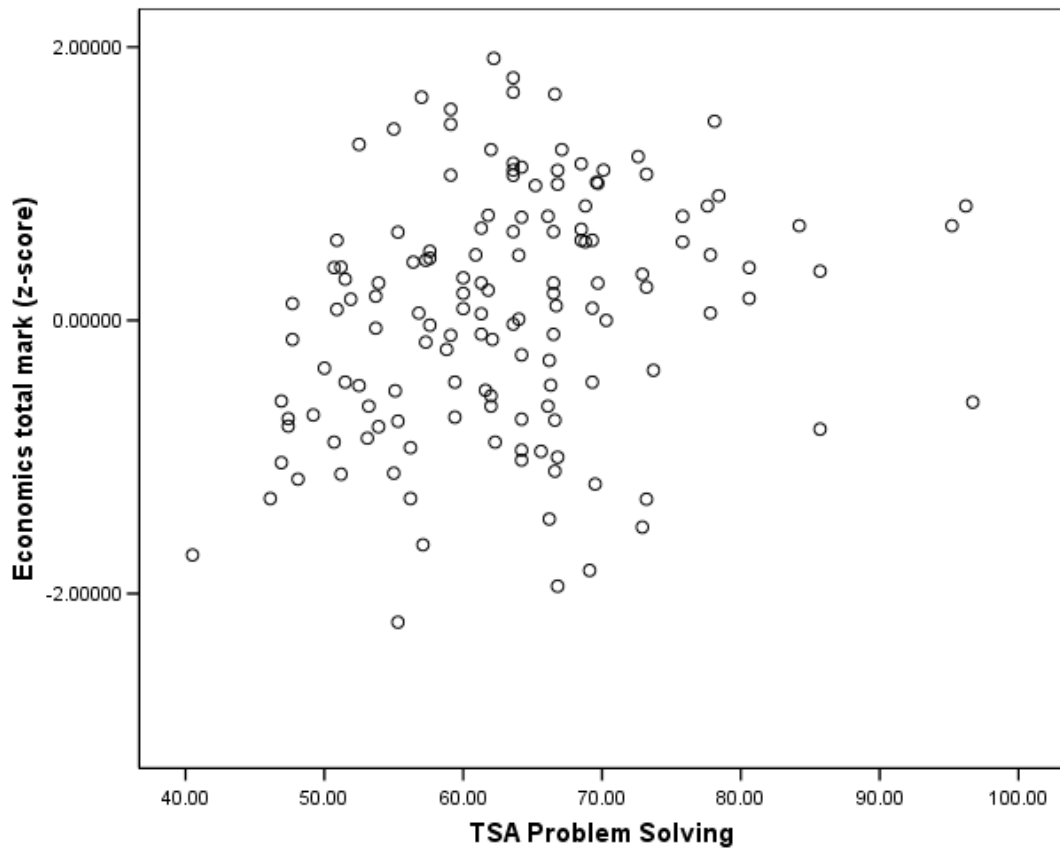
Figure 1: Scattergrams showing the relationship between TSA scores and 1<sup>st</sup> year examination performance

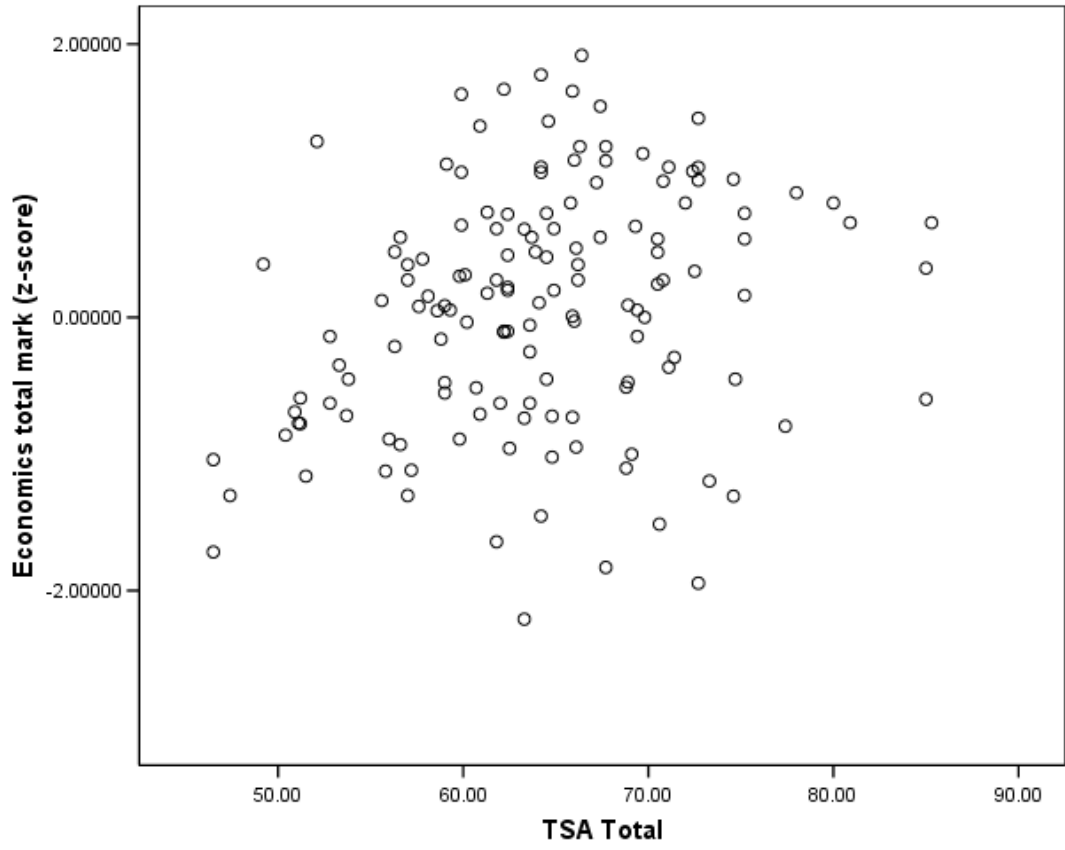
a) Computer Science



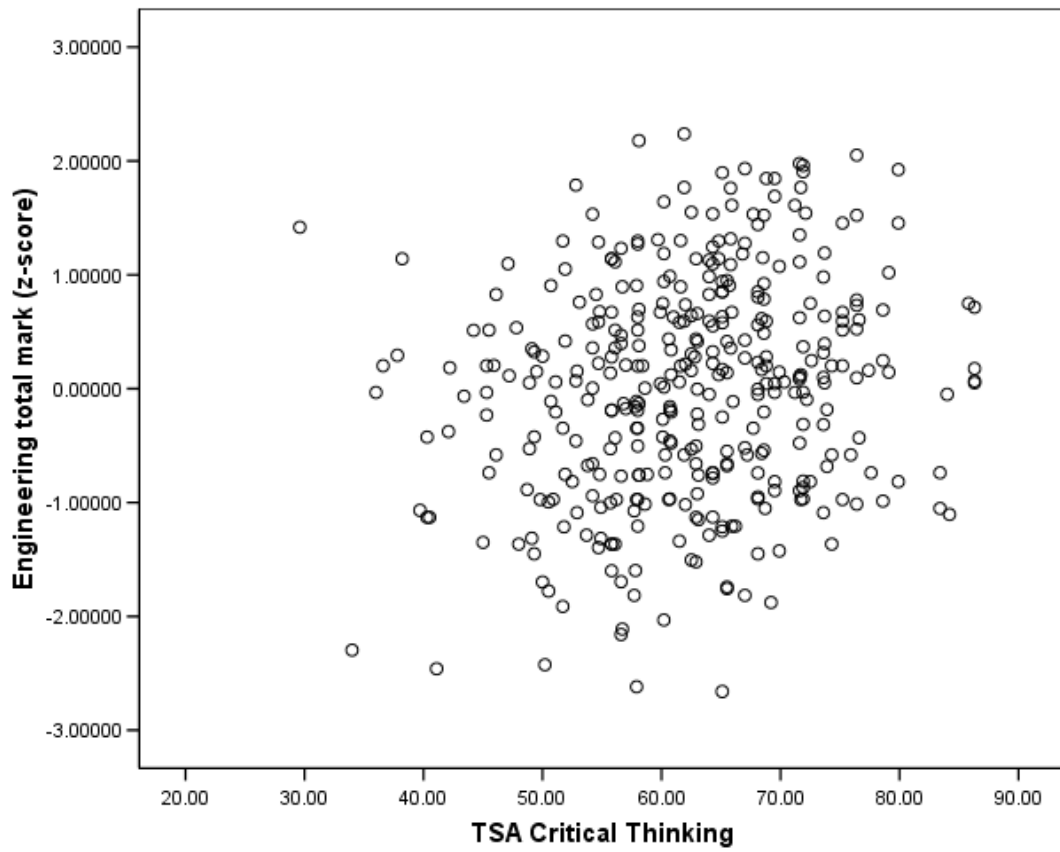
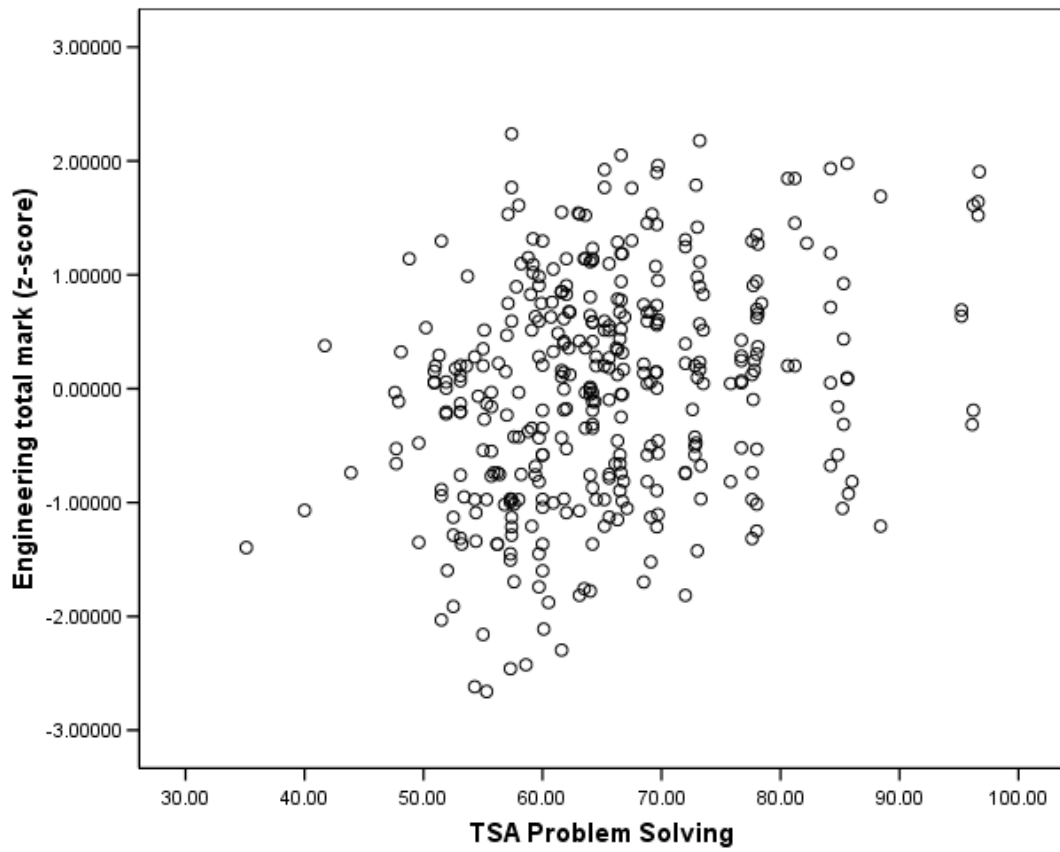


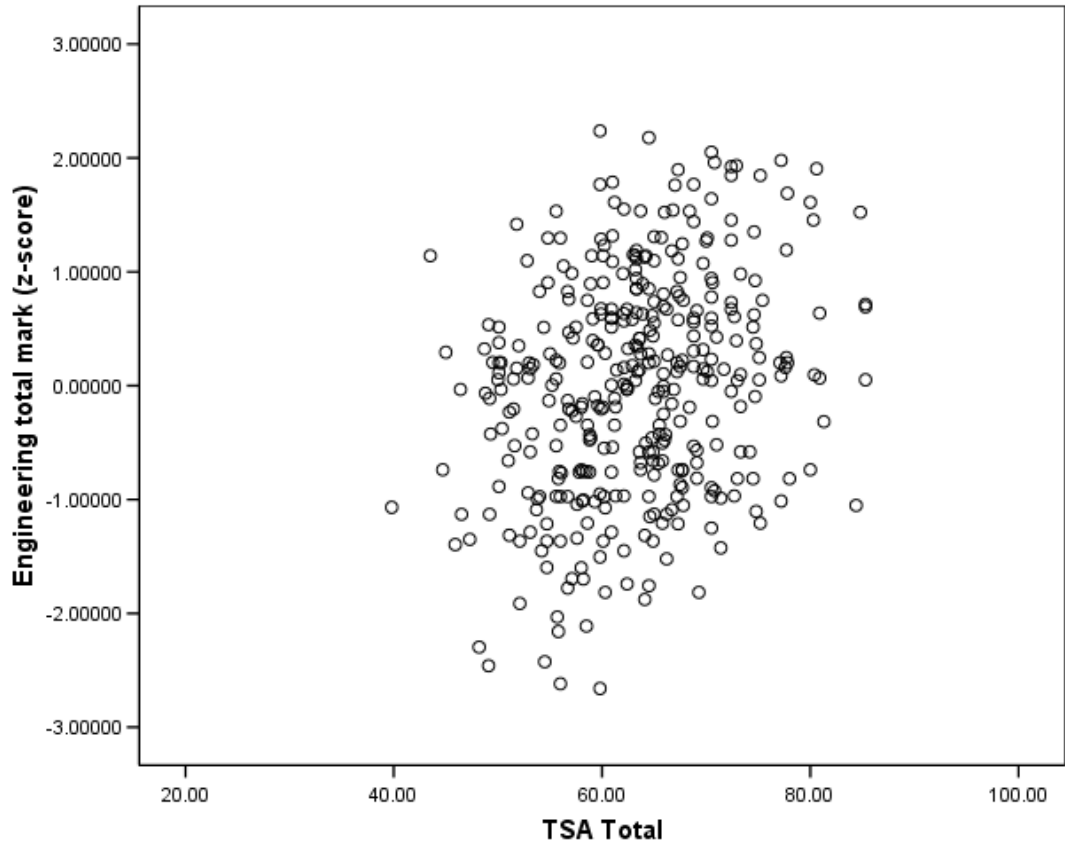
b) Economics



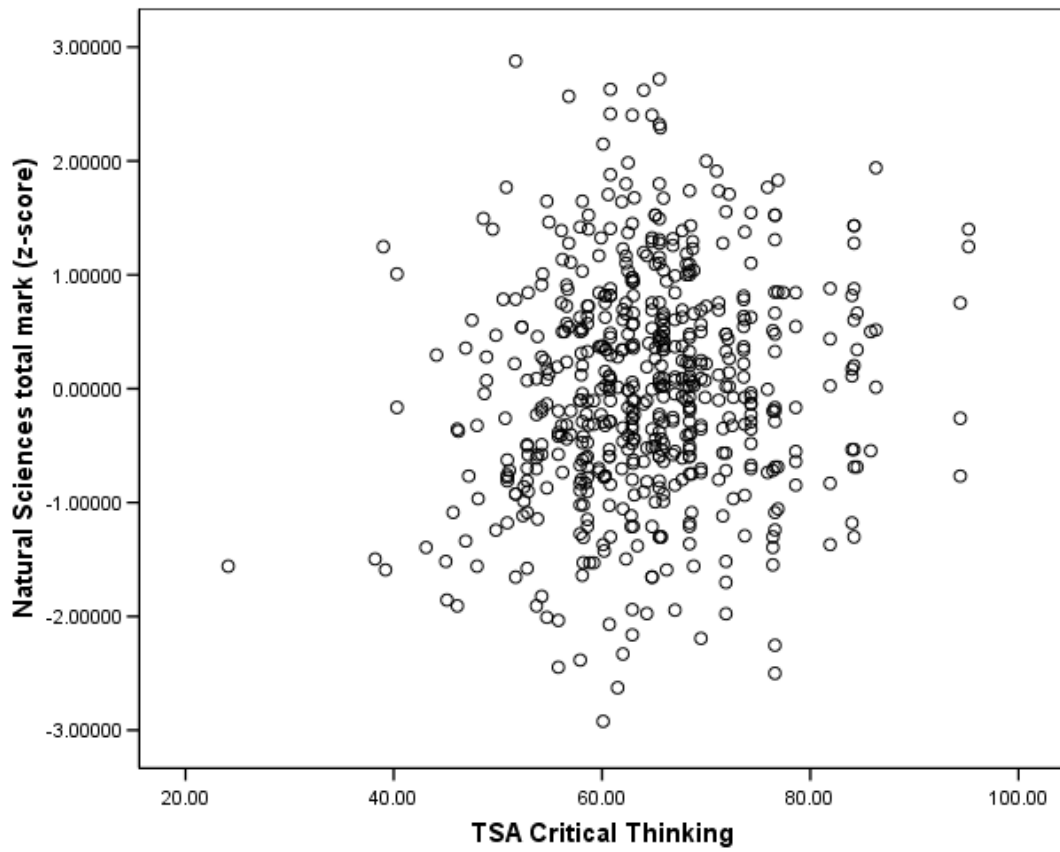
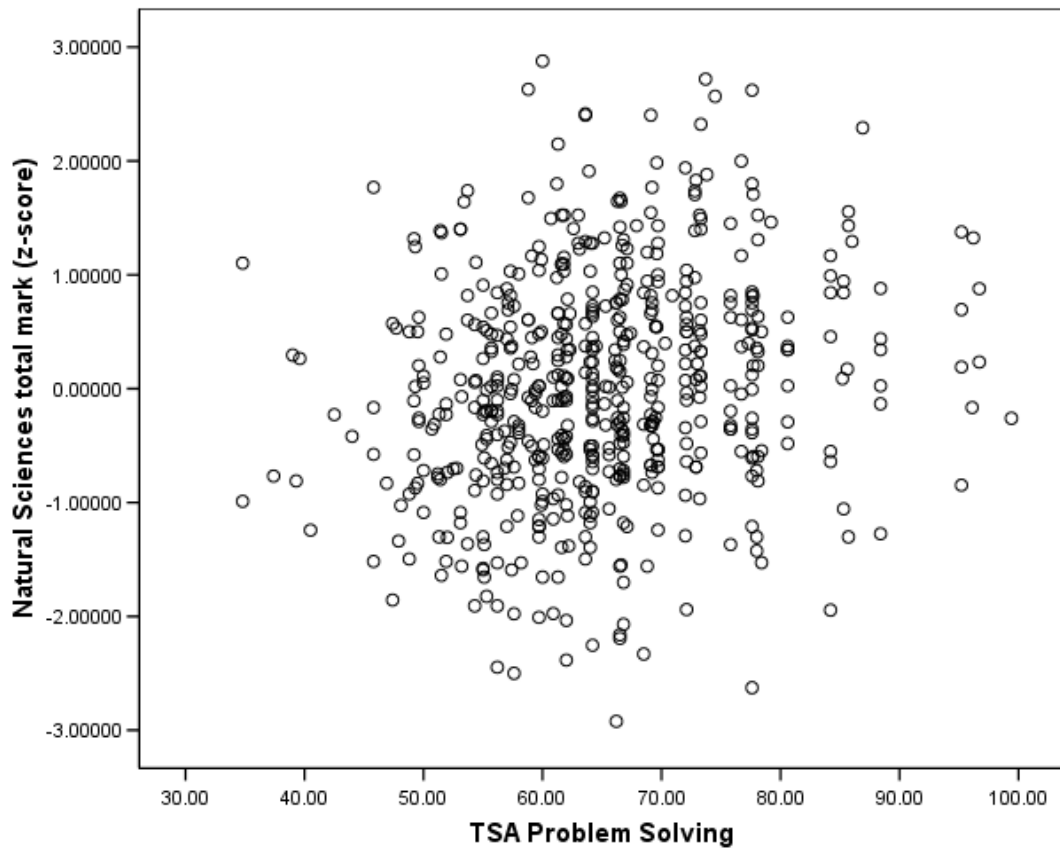


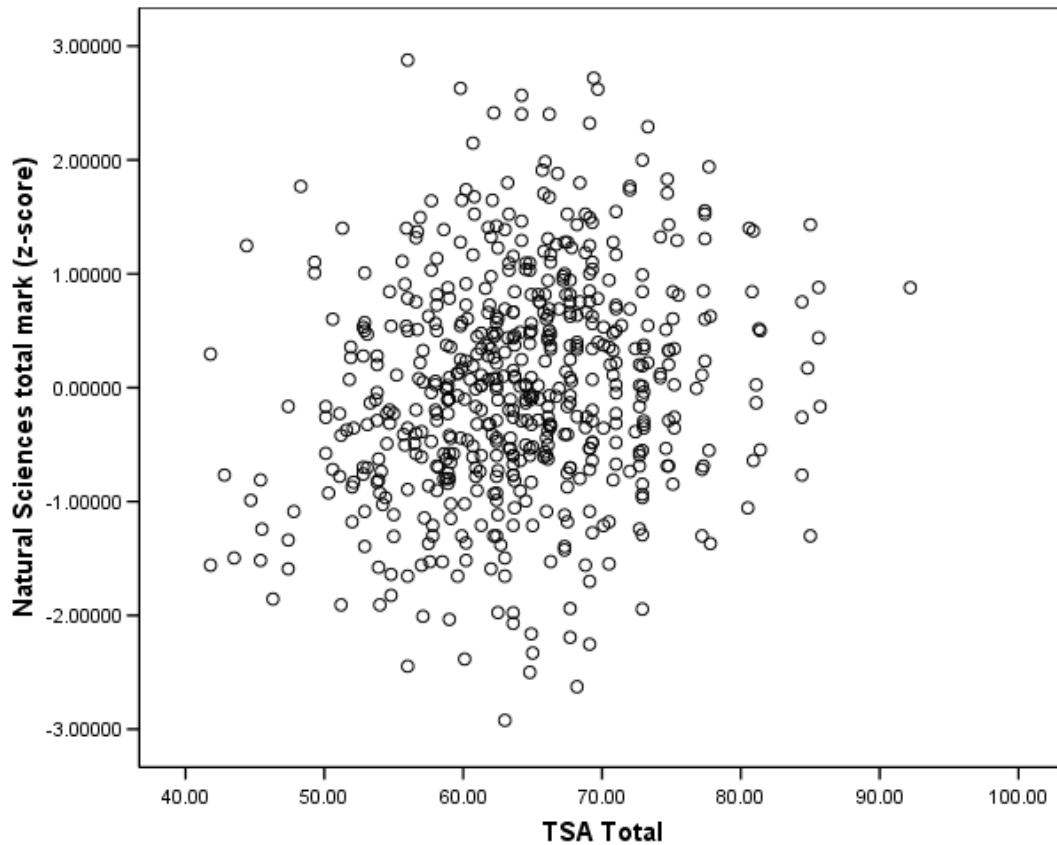
c) Engineering





d) Natural Sciences

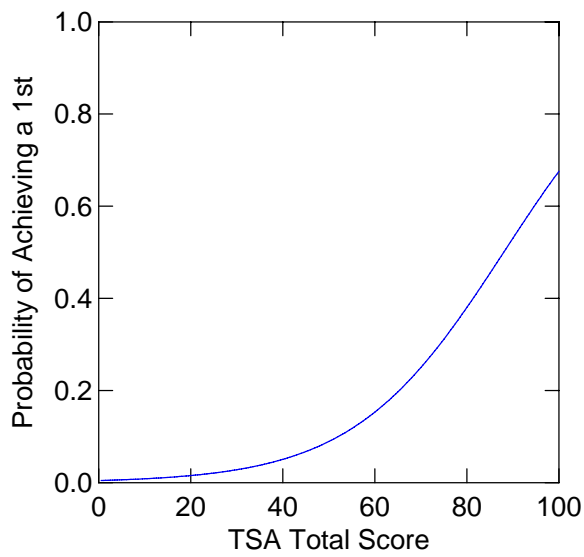
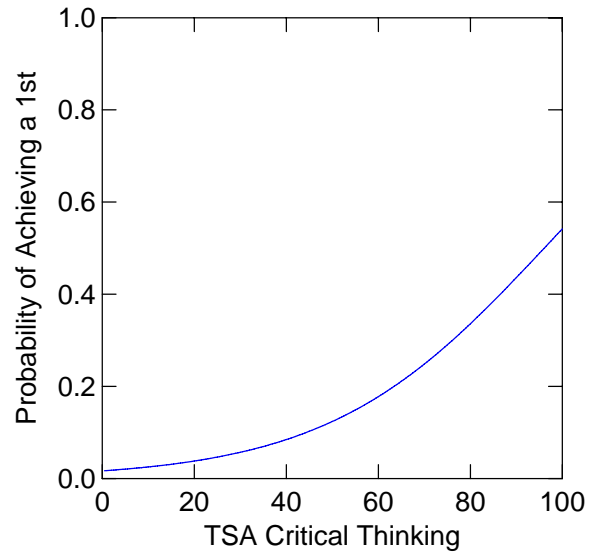
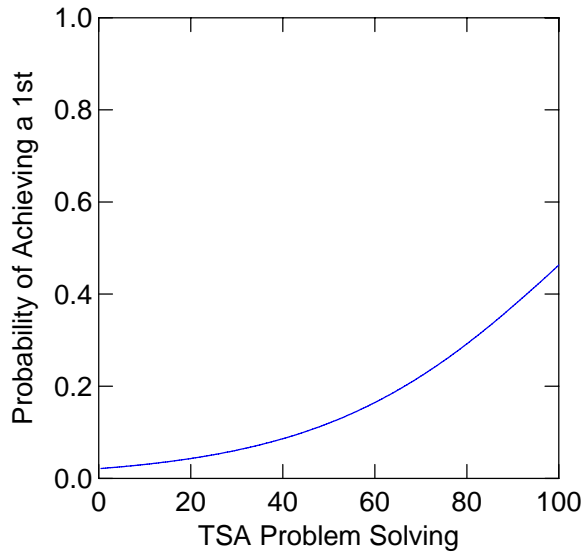




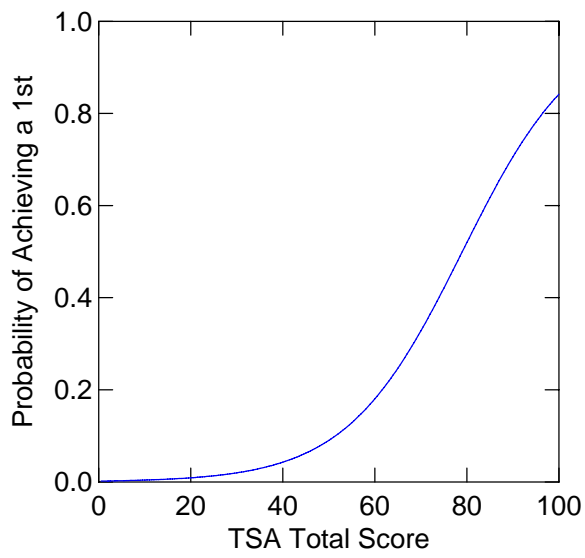
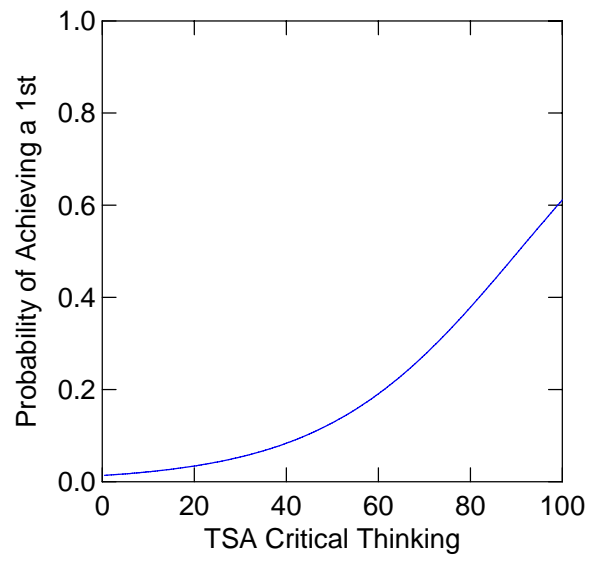
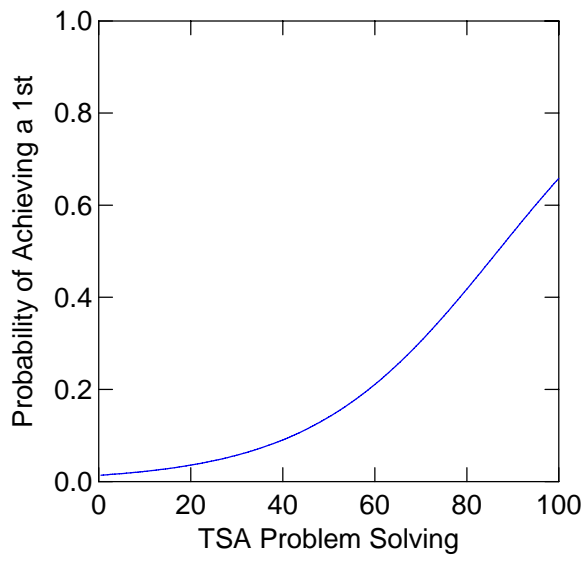
Correlations between TSA scores and overall 1<sup>st</sup> year examination performance fall mainly in the 'likely to be useful' range when the three cohorts are combined. All are significant at the 0.01 level. Coefficients are slightly higher for the Problem Solving component in all four courses. A previous report has shown that higher correlation coefficients are obtained if students for whom English is not their first language are analysed separately (see Emery & Shannon, 2007) and work is ongoing on this. The strength of the predictive relationships is hindered by the compensatory nature of the selection process (low TSA scorers who are admitted onto the courses are likely to have excelled in some other way that compensated for this, making them atypical of low scorers in general). Another problem is that the reasons for poor examination performance can be unrelated to academic ability.

Figure 2: Logistic regression plots showing the probability of achieving a 1<sup>st</sup> class outcome as a function of TSA scores

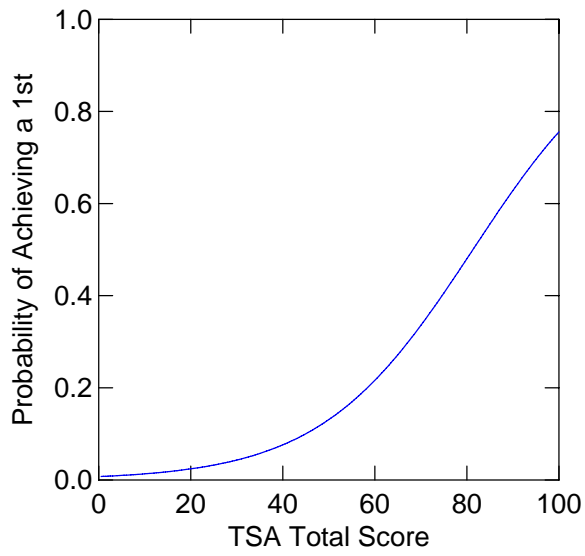
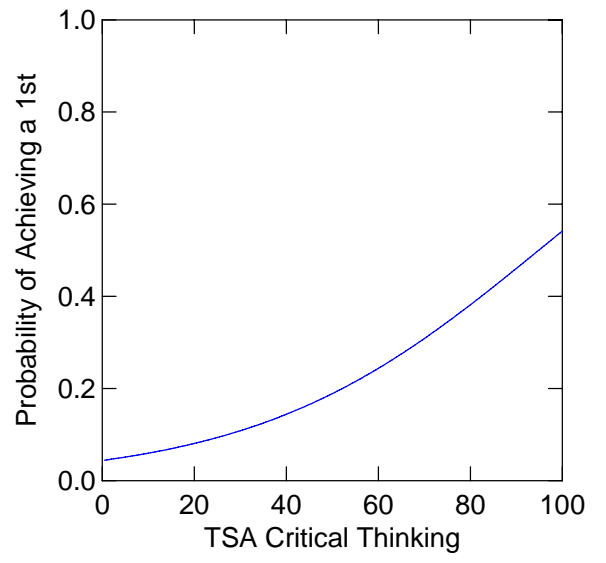
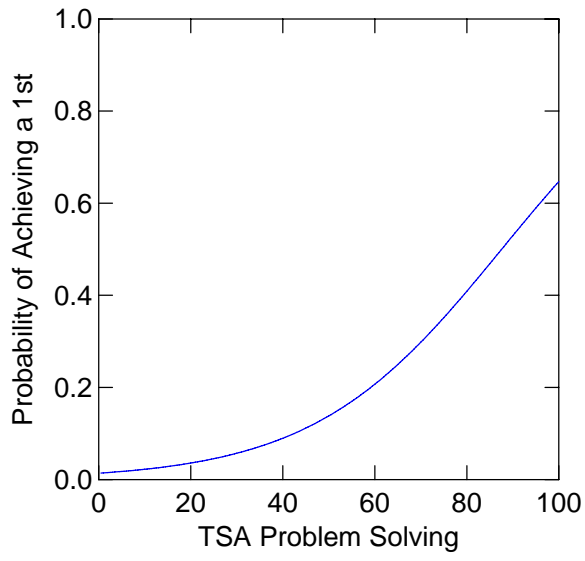
a) Computer Science (observations=172, 1sts=38)



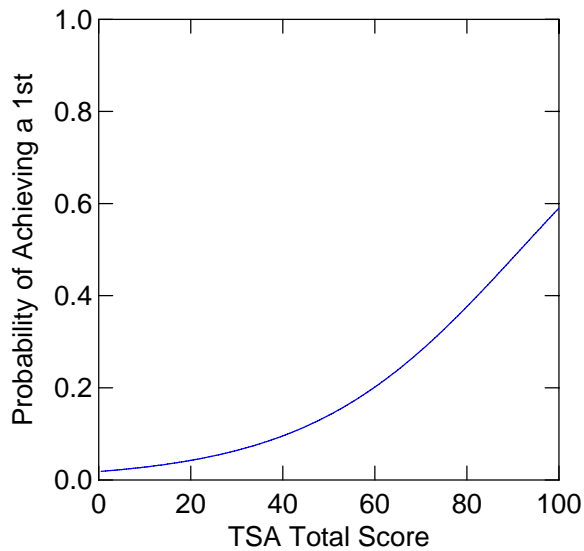
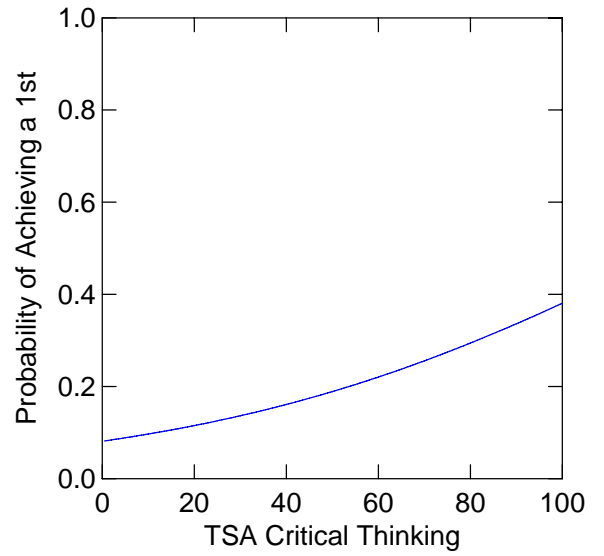
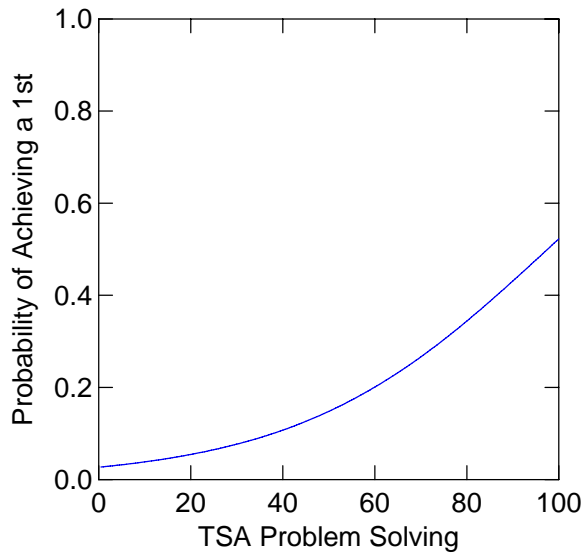
b) Economics (observations=145, 1sts=36)



c) Engineering (observations=370, 1sts=97)



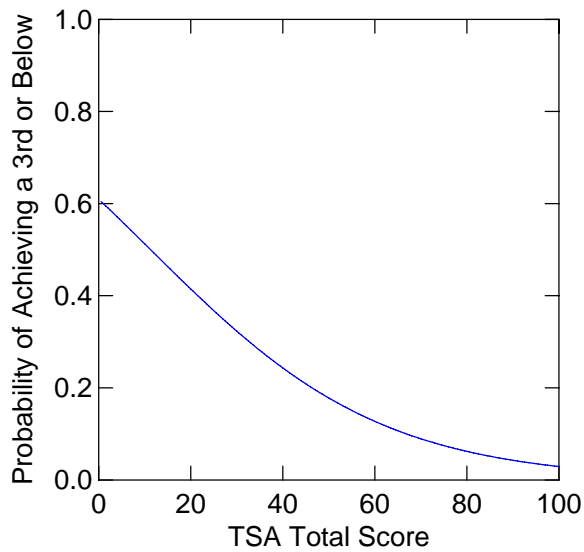
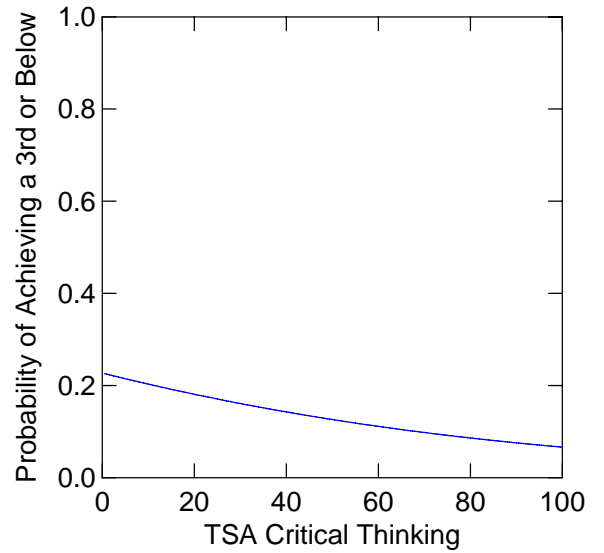
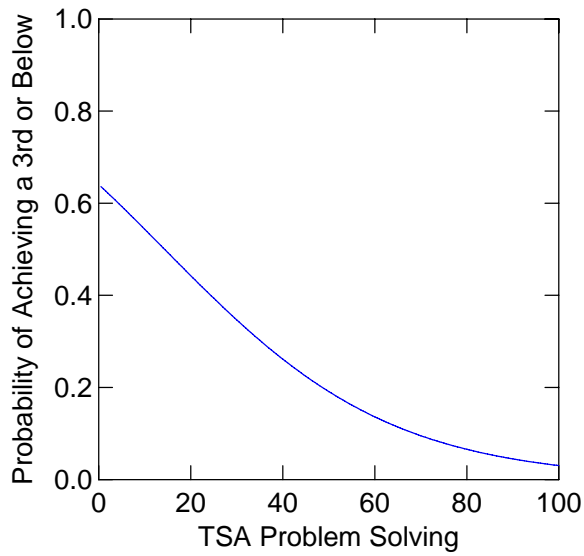
d) Natural Sciences (observations=577, 1sts=137)



The logistic regression plots in Figure 2 suggest that the combined total TSA score is the strongest predictor of achieving the highest examination class. An increase in scores here makes the greatest difference to the probability of success (the strength of the predictive relationship is directly implied by the vertical steepness of the curve in logistic regression). Plots should be interpreted with reference to the scattergrams in Figure 1, which show the actual ranges of TSA scores achieved by the accepted candidates.

Figure 3: Logistic regression plots showing the probability of achieving a 3<sup>rd</sup> class outcome or below as a function of TSA scores

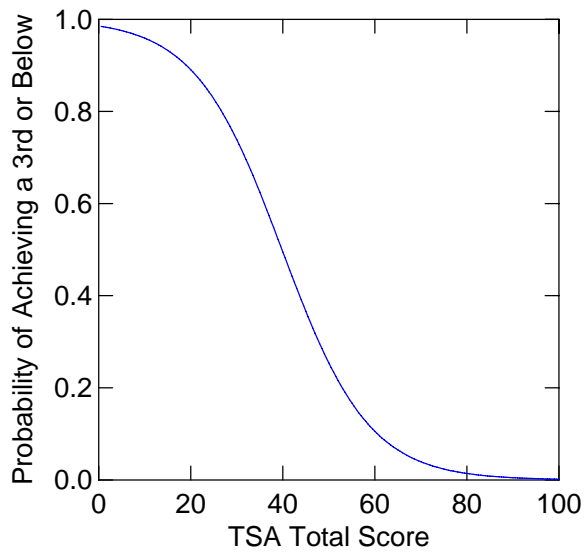
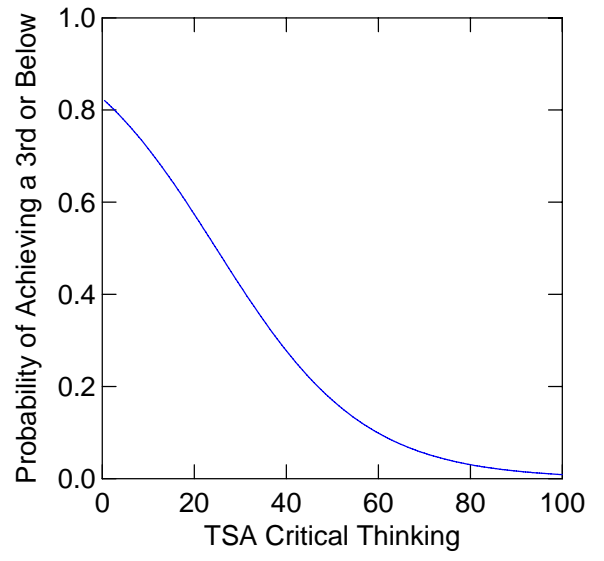
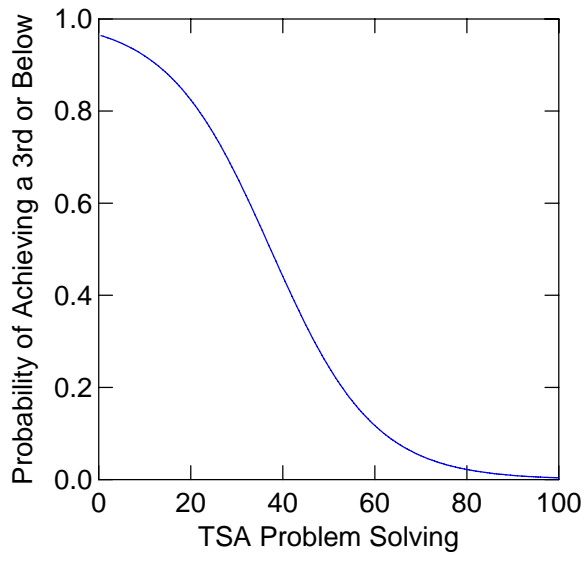
a) Computer Science (observations=172, 3rds or below=18)



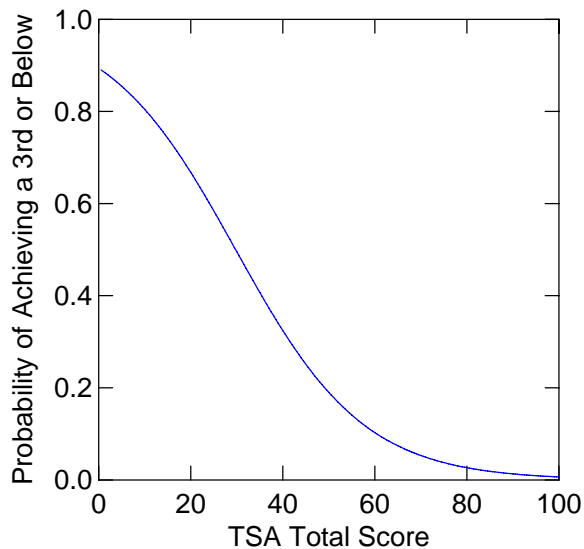
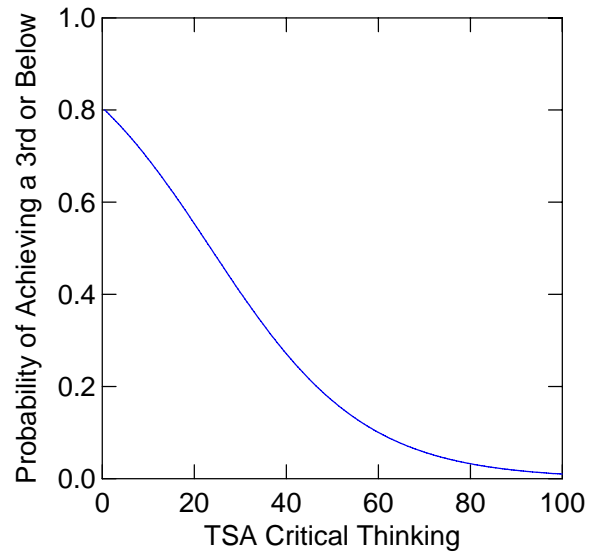
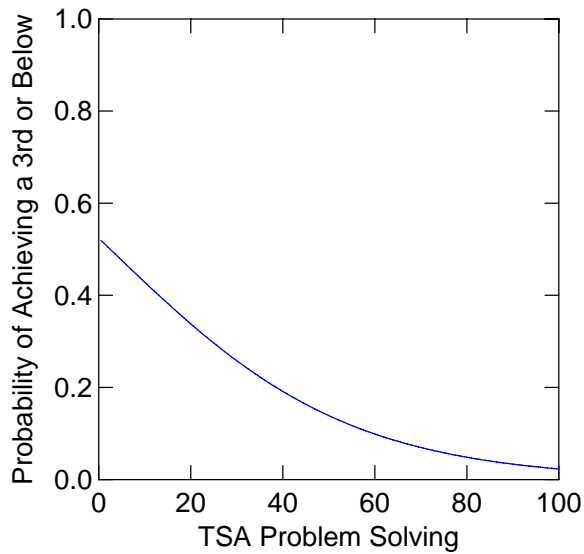
b) Economics (observations = 145, 3rds or below = 3)

Not enough cases for analysis

c) Engineering (observations=370, 3rds or below=37)



d) Natural Sciences (observations=577, 3rds or below=51)



The logistic regression plots in Figure 3 suggest that TSA scores predict the probability of achieving a 3<sup>rd</sup> class examination outcome or below, despite the popular notion that poor examination performance is often for reasons of a non-academic nature. Again, the combined total score makes the greatest difference to the probability of achieving this outcome. The relationships are particularly strong for the Engineering course.

For the Natural Sciences course, the TSA Problem Solving component appears to better predict the probability of achieving a 1<sup>st</sup> class outcome whilst the Critical Thinking component better predicts the probability of a 3<sup>rd</sup> class outcome or below. The reverse is true for Computer Science, where Critical Thinking scores better predict a high examination outcome and a poor examination outcome is more strongly associated with low Problem Solving scores.

## Conclusions

When three cohorts of students are analysed together, correlations with the TSA appear similar for Computer Science, Economics and Engineering and are slightly weaker for Natural Sciences. A possible reason for this is the diversity of the Natural Sciences course: total examination marks are composed of many possible combinations of options, some of which are predicted more strongly than others (see previous reports available on the TSA website). Logistic regression functions indicate that total TSA scores are a strong predictor of achieving both the highest and the lowest examination classes.

## References

Emery, JL, & Shannon, MD. Summary of TSA research. Available on the TSA website ~  
[http://www.tsa.ucles.org.uk/pdf/Summary%20of%20TSA%20research%20\(01-06-07\).pdf](http://www.tsa.ucles.org.uk/pdf/Summary%20of%20TSA%20research%20(01-06-07).pdf)

Sackett PR, Yang H. (2000) Correction for range restriction: an expanded typology. *Journal of Applied Psychology*, **85**:112-8.

US Department of Labor, Employment Training and Administration (1999) Employers' Guide to Good Practices. Accessed at [www.doleta.gov](http://www.doleta.gov)

The Cambridge Assessment Group is Europe's largest assessment agency. It plays a leading role in researching, developing and delivering assessment to eight million candidates in 150 countries.

Cambridge Assessment  
1 Hills Road  
Cambridge CB1 2EU  
United Kingdom  
tel +44 (0) 1223 553311  
fax +44 (0) 1223 460278



Cambridge Assessment is the brand name of the University of Cambridge Local Examinations Syndicate, a department of the University of Cambridge. Cambridge Assessment is a not-for-profit organisation.