Writing in tables and lists:
Exploring multimodal undergraduate writing through keyword searches

Maria Leedham
m.e.leedham@open.ac.uk

ICAME 2011
Outline

1. Research questions and the corpora

2. Findings
   2.1 Overall findings from keyword analysis
   2.2 Visuals and lists
   2.2 Comparison of text pairs in Biology and Economics

3. Conclusions
Why study UK undergraduate writing?

- UG assessed writing is a ‘high-stakes’ genre which has been under-researched

- UG writing is challenging due to the recent increase in multi-disciplinary degrees…
  **But**… academic writing varies between disciplines (e.g. Hewings, 1999; Hyland, 2008)

- Increase in new genres in UK assignments e.g. reflective blogs, website evaluations or press releases (Leedham, 2009; Nesi & Gardner, 2006).

- A major strategic aim of assignment-writing is to display disciplinary knowledge in an appropriate form
Why study Chinese students’ writing?

- The ‘largest single overseas student group in the UK’ (British Council, 2010)
  Over 85,000 Chinese students in UK in 2009

- **BUT**… most studies of Chinese students’ writing have been carried out on learner corpora (e.g. Chuang and Nesi, 2006; Mayor et al., 2007) or postgraduate theses (e.g. Hyland, 2008)

- This study uses authentic undergraduate assignments

Research Questions

1. What are some of the differences between Chinese and British undergraduate students' assessed UG writing?

2. How do these vary across the years of undergraduate study?

3. How significant is the discipline of study?
The Corpora

British Academic Written English (BAWE)
- 6,506,995 words
- 2,896 texts
- 2,761 assignments
- 1,039 contributors
- 30+ disciplines
- 13 genre families
- 4 levels of study
- Variety of L1s
- All proficient writing

The corpora for this study
- Extracted L1 English and L1 Chinese texts from BAWE
- Reduced to UG texts only
- Selected 5 disciplines
- Added extra L1 Chinese texts from other sources
- Resulting in: 104 texts from Chinese students
- 295 texts from British students
- Insights from lecturer interviews

ESRC project number
RES-000-23-0800
The corpora

<table>
<thead>
<tr>
<th>Discipline</th>
<th>L1 Chinese corpus</th>
<th></th>
<th></th>
<th>L1 English corpus</th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>No. of assignments</td>
<td>No. of words</td>
<td>Av. length</td>
<td>No. of assignments</td>
<td>No. of Words</td>
<td>Av. length</td>
</tr>
<tr>
<td>Biological Science</td>
<td>18</td>
<td>33,633</td>
<td>1868</td>
<td>83</td>
<td>173,412</td>
<td>2089</td>
</tr>
<tr>
<td>Business</td>
<td>20</td>
<td>33,303</td>
<td>1665</td>
<td>37</td>
<td>82,966</td>
<td>2242</td>
</tr>
<tr>
<td>Economics</td>
<td>20</td>
<td>38,086</td>
<td>1904</td>
<td>22</td>
<td>52,158</td>
<td>2371</td>
</tr>
<tr>
<td>Engineering</td>
<td>20</td>
<td>35,627</td>
<td>1781</td>
<td>97</td>
<td>203,782</td>
<td>2101</td>
</tr>
<tr>
<td>Food Science</td>
<td>26</td>
<td>30,267</td>
<td>1164</td>
<td>55</td>
<td>73,496</td>
<td>1336</td>
</tr>
<tr>
<td>Totals</td>
<td>104</td>
<td>170,916</td>
<td></td>
<td>294</td>
<td>585,814</td>
<td></td>
</tr>
</tbody>
</table>
Outline

1. Research questions and the corpora

2. Findings
   2.1 Overall findings from keyword analysis
   2.2 Visuals and lists
   2.2 Comparison of text pairs in Biology and Economics

3. Conclusions
Keyness

• ‘A word which is *positively* key occurs *more* often than would be expected by chance in comparison with the reference corpus.’
  Scott, WordSmith Help files, 2010

• WordSmith Tools v.5 (Scott, 2010)

• Used log likelihood statistic, $p=0.000001$

• Extracted key words and key n-grams (2-5 words)

• Compared keywords in Chi-Engineering with Eng-Engineering, and in each corpus with all-UG-BAWE, etc

• Checked to ensure each keyword or key n-gram occurs in writing from at least 5 texts and 3 students
### Keywords

<table>
<thead>
<tr>
<th>Key in Chi-Economics</th>
<th>Rate, model, output, formula, level, growth, curve, income, government, supply, students, population, dividends, per, reserves, consumption, T, aggregate, tax, Dutch, quantity, stock, portfolio, assets, inefficiency, competitive, capm, generation, repurchases, qtmark, asset, refer, cash, disposable, progress, deficit, income</th>
</tr>
</thead>
<tbody>
<tr>
<td>Key in Eng-Economics</td>
<td>In, that, as market, however, if, policy, economy, firm, therefore, firms, had, competition, costs, hence, under, since, U.S., significant, period, shown, war, international, lower, world, did, Britain, markets, impact, profits, transport, Bertrand, railways, crises, states, you, cournot, wages, question, extent, stabilisation, British, vertical, shirking, credibility, IMF, governments</td>
</tr>
<tr>
<td>Key in both</td>
<td>Price, demand, monopoly, well, than, capital, increase, higher, exchange, inflation, labour, economic, unemployment, prices, countries, money, production, cost, investment, interest, firm, foreign, crisis, trade, wage, long, marginal, F, country, Y, run, elasticity, domestic, variables, goods, exam, equilibrium, expectations, rates, short, consumers, monetary, surplus, policies, consumer, efficiency, spending, scale, fiscal, productivity, Phillips, slope, bank, central, monopolist, saving, relative</td>
</tr>
</tbody>
</table>

- first person pronouns (we, I)
- connectors
- references to tables and figures
- use of numbers in lists (denoted in WS by # in some disciplines)
First person pronouns

<table>
<thead>
<tr>
<th></th>
<th>Chi Biol</th>
<th>Eng Biol</th>
<th>Chi Engin</th>
<th>Eng Engin</th>
<th>Chi Econ</th>
<th>Eng Econ</th>
</tr>
</thead>
<tbody>
<tr>
<td>I</td>
<td>1</td>
<td>1</td>
<td>3</td>
<td>10****</td>
<td>9</td>
<td>14*</td>
</tr>
<tr>
<td>we</td>
<td>6</td>
<td>7</td>
<td>17</td>
<td>15</td>
<td>29</td>
<td>23</td>
</tr>
<tr>
<td>Total</td>
<td>7</td>
<td>8</td>
<td>20</td>
<td>25</td>
<td>38</td>
<td>37</td>
</tr>
</tbody>
</table>

* p<.05; ** p<.01; *** p<.001; **** p<.0001

Use of I in reflective writing in Engineering

(1) … I don't think this is what a professional engineer is, although I do think that a professional engineer must work in this ‘professional manner’ … (0354f).

(2) With hindsight I would have called everybody an hour before the meeting to make sure they were coming as I eventually had to later on in the project (0342c).
## Connectors

<table>
<thead>
<tr>
<th></th>
<th>Chi-Economics (raw)</th>
<th>Eng-Economics (raw)</th>
</tr>
</thead>
<tbody>
<tr>
<td>however</td>
<td>14 (55)</td>
<td>34**** (178)</td>
</tr>
<tr>
<td>therefore</td>
<td>17 (65)</td>
<td>23* (122)</td>
</tr>
<tr>
<td>hence</td>
<td>8 (32)</td>
<td>16*** (86)</td>
</tr>
<tr>
<td>thus</td>
<td>12**** (44)</td>
<td>4 (20)</td>
</tr>
<tr>
<td>in contrast</td>
<td>2** (7)</td>
<td>0 (1)</td>
</tr>
</tbody>
</table>

* p<.05; ** p<.01; *** p<.001; **** p<.0001

Preferred connectors… and difference in position

(3) …about the prices of other firms. However, monopolies can charge what they like. (6008q, L1 chinese)

(4) The implication of this however, is that price discrimination, which is possible through monopoly… (0399a L1 English)
Outline

1. Research questions and the corpora

2. Findings
   2.1 Overall findings from keyword analysis
   2.2 **Visuals and lists**
   2.2 Comparison of text pairs in Biology and Economics

3. Conclusions
### Keywords

<table>
<thead>
<tr>
<th>Key in Chi-Economics</th>
<th>rate, model, output, formula, level, growth, curve, income, government, supply, students, population, dividends, per, reserves, consumption, T, aggregate, tax, Dutch, quantity, stock, portfolio, assets, inefficiency, competitive, capm, generation, repurchases, qtmark, asset, refer, cash, disposable, progress, deficit, income</th>
</tr>
</thead>
<tbody>
<tr>
<td>Key in Eng-Economics</td>
<td>in, that, as market, however, if, policy, economy, firm, therefore, firms, had, competition, costs, hence, under, since, U.S., significant, period, shown, war, international, lower, world, did, Britain, markets, impact, profits, transport, Bertrand, railways, crises, states, you, cournot, wages, question, extent, stabilisation, British, vertical, shirking, credibility, IMF, governments</td>
</tr>
<tr>
<td>Key in both</td>
<td>price, demand, monopoly, we, than, capital, increase, higher, exchange, inflation, labour, economic, unemployment, prices, countries, money, production, cost, investment, interest, firm, foreign, crisis, trade, wage, long, marginal, F, country, Y, run, elasticity, domestic, variables, goods, exam, equilibrium, expectations, rates, short, consumers, monetary, surplus, policies, consumer, efficiency, spending, scale, fiscal, productivity, Phillips, slope, bank, central, monopolist, saving, relative</td>
</tr>
</tbody>
</table>

- first person pronouns (we, I)
- connectors
- references to tables and figures
- use of numbers in lists (denoted in WS by # in some disciplines)
**Visuals and lists**

**Visuals**: tables, figures, diagrams, images, photos...

**Lists**
- A ‘genuine’ or prototypical list contains ‘list items’, each consisting of a word or NP/VP.
- List items are separated by bullet points/hyphens/letters/numbers, or are indented.

**Listlikes**
- ‘paragraphs of running text carrying list-like formatting’
- ‘false lists’ or ‘list-likes’ contain larger units of text per list item (Heuboeck et al., 2005: 29).

<table>
<thead>
<tr>
<th>Carbon Content %</th>
<th>Classification</th>
<th>Uses</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.3-0.4</td>
<td>Low Carbon Steel</td>
<td>General purpose steel used for welding. Poor corrosion resistance</td>
</tr>
<tr>
<td>0.3-0.7</td>
<td>Medium Carbon Steel</td>
<td>Used for the production of nuts, shafts, and gears. Very difficult to weld</td>
</tr>
<tr>
<td>0.7-1.7</td>
<td>High Carbon Steel</td>
<td>Used in high stress applications, such as springs, and as cutting tools</td>
</tr>
</tbody>
</table>

Table 1
Visuals and lists

Lists

A ‘genuine’ or prototypical list contains ‘list items’, each consisting of a word or NP/VP. List items are separated by bullet points/hyphens/letters/numbers, or are indented.

Listlikes

- ‘paragraphs of running text carrying list-like formatting’
- ‘false lists’ or ‘list-likes’ contain larger units of text per list item (Heuboeck et al., 2005: 29).

Conclusions

The experiment yielded the following conclusions:
- The efficiency of a single stage centrifugal pump at high pump speed (3000 RPM) is better than it at low pump speed (2000 RPM).
- The input power with high pump speed increases faster than the one with low pump speed as discharge increases.
- The relationship between total head and discharge is not affected by pump speed, but higher pump speed provides higher total head.
## Keywords relating to visuals and lists

<table>
<thead>
<tr>
<th>L1&amp; discipline</th>
<th>Chi-Biol</th>
<th>Chi-Bus</th>
<th>Chi-Econ</th>
<th>Chi-Engin</th>
<th>Chi-Food</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Selected keywords</strong></td>
<td>#</td>
<td>growth</td>
<td>#</td>
<td>#</td>
<td>#</td>
</tr>
<tr>
<td>table</td>
<td>curve</td>
<td>eq.</td>
<td>curve</td>
<td></td>
<td></td>
</tr>
<tr>
<td>data</td>
<td>refer</td>
<td>according</td>
<td>referring</td>
<td></td>
<td></td>
</tr>
<tr>
<td>equation</td>
<td>model</td>
<td>figure</td>
<td>statistical</td>
<td></td>
<td></td>
</tr>
<tr>
<td>figure</td>
<td>per</td>
<td></td>
<td>deviation</td>
<td></td>
<td></td>
</tr>
<tr>
<td>graph</td>
<td>output</td>
<td></td>
<td>numbers</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
References to visuals and lists

Biology
• All "Phases" are labeled on the graph. The curve of the Exponential Phase was straight, though some point lay outside this best straight line of fit (0041a).

Economics
• Actually the total loss resulting from the lower monopoly output (Q M) is the grey triangle. The part of the grey triangle above P C is the loss of consumer surplus (6008q).

Engineering
• According to the program and refer to the figure 4.1.1, it is easy to find… (6107d).

Food Science
• According to the 3 sets of data calculated above… (6150d).
References to visuals and lists

Biology
• All "Phases" are labeled on the graph. The curve of the Exponential Phase was straight, though some point lay outside this best straight line of fit (0041a).

Economics
• Actually the total loss resulting from the lower monopoly output (Q M) is the grey triangle. The part of the grey triangle above P C is the loss of consumer surplus (6008q).

Engineering
• According to the program and refer to the figure 4.1.1, it is easy to find… (6107d).

Food Science
• According to the 3 sets of data calculated above… (6150d).
# Visuals and lists

<table>
<thead>
<tr>
<th></th>
<th>Tables</th>
<th>Figures</th>
<th>Lists</th>
<th>Listlikes</th>
</tr>
</thead>
<tbody>
<tr>
<td>Chi-Biol</td>
<td>15****</td>
<td>25****</td>
<td>1</td>
<td>4</td>
</tr>
<tr>
<td>Eng-Biol</td>
<td>5</td>
<td>13</td>
<td>2</td>
<td>6</td>
</tr>
<tr>
<td>Chi-Econ</td>
<td>1</td>
<td>14****</td>
<td>2*</td>
<td>25****</td>
</tr>
<tr>
<td>Eng-Econ</td>
<td>0</td>
<td>12</td>
<td>1</td>
<td>3</td>
</tr>
<tr>
<td>Chi-Bus</td>
<td>2</td>
<td>2</td>
<td>6*</td>
<td>129****</td>
</tr>
<tr>
<td>Eng-Bus</td>
<td>6**</td>
<td>6**</td>
<td>3</td>
<td>23</td>
</tr>
<tr>
<td>Chi-Food</td>
<td>20*</td>
<td>6</td>
<td>5</td>
<td>82****</td>
</tr>
<tr>
<td>Eng-Food</td>
<td>14</td>
<td>6</td>
<td>4</td>
<td>18</td>
</tr>
<tr>
<td>Chi-Engin</td>
<td>10*</td>
<td>21</td>
<td>7</td>
<td>53****</td>
</tr>
<tr>
<td>Eng-Engin</td>
<td>7</td>
<td>21</td>
<td>10</td>
<td>24</td>
</tr>
</tbody>
</table>

* p<.05
** p<.01
**** p<.0001
Listlikes by yeargroup

![Bar chart showing listlikes by yeargroup subcorpora. Chi3 has the highest count, followed by Chi2, while Eng1, Eng2, and Eng3 have comparatively lower counts.](chart.png)
Outline

1. Research questions and the corpora

2. Findings
   2.1 Overall findings from keyword analysis
   2.2 Visuals and lists
   2.2 Comparison of text pairs in Biology and Economics

3. Conclusions
Visuals and extended captions in Biology

• Dinolfo et al. (2007) students ‘see’ or ‘read’ cells under a microscope and subsequently describe them.

• “all-at-once’ processing of complex and often competing visual data’ compared with the ‘linear ‘one-at-a-time’ processing that occurs when we read written text line by line’ (Dinolfo et al., 2007: 401)

Comparison of two Biology assignments

<table>
<thead>
<tr>
<th>Text feature</th>
<th>Chinese, text 0434a</th>
<th>English, text 0067b</th>
</tr>
</thead>
<tbody>
<tr>
<td>No. of pages excluding refs</td>
<td>15.5</td>
<td>9</td>
</tr>
<tr>
<td>No. of tokens (in WS)</td>
<td>3234</td>
<td>3201</td>
</tr>
<tr>
<td>No. of tables</td>
<td>2</td>
<td>0</td>
</tr>
<tr>
<td>No. of figures</td>
<td>17</td>
<td>5</td>
</tr>
<tr>
<td>Visuals as proportion of whole text</td>
<td>48% (7.5 pp)</td>
<td>22% (2pp)</td>
</tr>
<tr>
<td>Layout</td>
<td>whole page</td>
<td>2 columns</td>
</tr>
</tbody>
</table>

No. of pages excluding refs 15.5 9 
No. of tokens (in WS) 3234 3201 
No. of tables 2 0 
No. of figures 17 5 
Visuals as proportion of whole text 48% (7.5 pp) 22% (2pp) 
Layout whole page 2 columns
The role of maternal effect genes in the development of the nematode *Caenorhabditis elegans*

**ABSTRACT**

*Caenorhabditis elegans* (C. elegans) has been used as one of the favourite model organisms for developmental studies. Embryogenesis of *C. elegans* extensively relies on maternal effect genes for intrinsically asymmetric cell division and cell-cell interactions. In this review, the early embryogenesis of *C. elegans*, from the establishment of Anterior-Posterior polarity initiated by sperm entry to the asymmetrical cell divisions and different cell lineages induced by a variety of cell fate determinant is summarized, some of the molecular mechanisms carried out by the crucial maternally expressed cell fate determinants underlying these processes are described.

**INTRODUCTION**

The *C. elegans* and its life cycle

*Caenorhabditis elegans* (C. elegans) is a small (~1mm long) free living soil nematode that has a predominantly hermaphroditic adult life. (Figure 1)

![Diagram of C. elegans](image)

Figure 1 Adult *C. elegans*. Upper diagram: differential interference contrast image of an adult *C. elegans*. Lower diagram: anatomical structures of adult *C. elegans* (schematic drawing). Middle Left scale bar: 0.1 mm

The life cycle of *C. elegans* contains an embryonic stage, four larval stages (L1-L4) and an adult stage. (Figure 2) Molt (apoplyosis, new cuticle formation, and ecdysis) takes place at the end of each larval stage. Under certain external conditions such as starvation, a non-growing stage, dauer larva, may form through a facultative, reversible, arrest at the lethargus in the second of four cuticle molts. The life cycle is about 2 to 3 weeks. Each approximately 8,000 *C. elegans* proteins have already been matched to homologous human gene transcripts (Lai et al., 2000). Specific mutations may be produced by targeted deletion through transposon insertion or mutagens. Embryos may be manipulated by transformation or injection with transgenes and marker proteins such as green fluorescent protein (GFP) are easily visualised in the transparent embryos. RNA interference (RNAi) is a particularly useful technique for studying maternal effect genes by eliminating the expression of specific maternal or zygotic genes in offspring.

**Reproduction**

In hermaphroditic worms, fertilization occurs in the spermatheca – an organ where the sperm is stored – when mature oocytes pass from the ovary towards the vulva (Fig 1A-B). The point of sperm entry determines the posterior end of the embryo. After fertilization, a rigid, ovoid-shaped chitin eggshell called the chorion is made (Kemphues & Strome, 1997) and the long axis of this oovid is termed the anteroposterior (ap) axis of the embryo.
Diagrams and extended caption in 0434a (Chinese writer)

- 186 word extended caption
- different font to main text
- text wrapping of image and caption
- full sentences and in same, neutral stance as the rest of the text (e.g. use of passives, no first person pronouns, formal language)
- caption describes the process illustrated by diagram on the left
- diagram and main text are not integrated (in this case) – the extended caption functions as a freestanding text

(Cf work on multimodal texts by Kress & Van Leeuwen)
Bulleted lists vs. connected prose in Economics

<table>
<thead>
<tr>
<th>Text feature</th>
<th>Chinese, 0155a</th>
<th>English, 0202j</th>
</tr>
</thead>
<tbody>
<tr>
<td>No. of pages excluding refs</td>
<td>6</td>
<td>6</td>
</tr>
<tr>
<td>No. of tokens (in WS)</td>
<td>3731</td>
<td>4242</td>
</tr>
<tr>
<td>No. of formulae</td>
<td>19</td>
<td>6</td>
</tr>
<tr>
<td>No. of lists</td>
<td>2</td>
<td>0</td>
</tr>
<tr>
<td>No. of listlikes</td>
<td>28</td>
<td>0</td>
</tr>
<tr>
<td>Lists and listlikes as % of whole text</td>
<td>90%</td>
<td>0%</td>
</tr>
</tbody>
</table>

These qualitative variables would have some impacts on the QMARK, but whether they are statistically significant will be investigated later.

Correlation matrices:
- It was found that QMARK has strong positive relationships with variables ABILITY, ALEVELSA, ATT, ATTL and strong negative relationships with variables EXPOLC and TOPB.
- The correlations between ATT, ATTL and ATTR are very high. Therefore, multicollinearity is an issue needed to be taken into account.
- EXPOLC has strong negative relationships with ATT, ATTL and ATTR and strong positive relationships with TOPB. It could be explained that students who got drunk could not get up easily to attend the lectures and classes. The more TOPB they attended, the more money they spent on alcohol.
- Generally, HRSQT has strong positive relationship with ALEVELSA and ALEVELS, while strong negative relationship with ABILITY and TOPB. This could be explained that the more TOPB students spent for, the fewer hours they spent on studying and the highest ability a student has the lesst hour of study is required for him/her. On the other hand, students having a good A level record maintained their hard working attitude.

2. Bivariate Regression and Multivariate Regression
(a) Bivariate Regression
1. The following results are obtained after running the bivariate regressions in EViews:
   - QMARK = 64.7070-0.002211ATTR

   Interpretation for the regression results:
     - The intercept 64.97 means that even students did not attend any revision lecture, they could get 64.97 in the exam which may not make much economic sense as revision lectures are designed to boost a student's exam marks.
     - Since coefficient of -0.0022 shows an inverse relationship between the attendance of revision lecture and exam performance. It means that 1% increase in the proportion of revision lecture attendance would decrease students' mark by 0.0022 in the exam.
     - Economic interpretation of this could be that students who attended revision lectures would spend more time revising topics mentioned in the revision lecture and ignoring other topics. However as the coefficient is small, we could hardly omit its effect.

   - R-squared value of about 0.000002 means that only 0.0002% of the variation in QMARK is explained by ATTR. Therefore it could be concluded that ATTR has such a trivial effect on exam performance that it could even be omitted.

Two-tailed t-test for the significance of the slope β:
- H0: β = 0 (Proportion of revision lecture attended does not affect exam performance)
- H1: H0 = 0 (Proportion of revision lecture attended does affect exam performance)
Since the calculated t-value -0.013 is lower than the critical value of t-test at 5% significance level with 370 d. f., we fail to reject H0 in this case and the conclusion is that revision lecture attendance does not affect exam performance.

(b) Multivariate Regression
   - QMARK = α + β1ATTR + β2ABILITY + β3HRSQT + u

   Modelling by GLS we get:
   - QMARK = 36.5252 + 0.105494ATTR + 0.804990ABILITY - 0.147658HRSQT

   Interpretation of the regression results:
   - α = 36.5252
   - β1 = 0.105494
   - β2 = 0.804990
   - β3 = -0.147658

   As reported in question 1, the correlation coefficient was 0.67.

Question 2b
Interpretation of results (equations 4 and 5 appendix 2)

- The coefficient on class attendance is 0.15, which implies that holding all other variables constant, if you increase class attendance by 1 unit (1% increase in class attendance in a year), then the exam mark will increase by 0.15 units (0.15% increase in your mark).
- The coefficient on lecture attendance is 0.06, meaning holding all other variables constant, attending 1% more lectures will increase your mark by 0.06%.
- The coefficient on revision lecture attendance is slightly surprising, at -0.04, implying that attending 1% more revision lectures, your mark will decrease by 0.04%. The intercept can be interpreted to mean that if you attended no classes, revisions or standard lectures, you would score 49.32%

Tests (shown in appendix 2)

- The coefficient on class attendance was significant at the 0.01 level implying that in the multiple regression model, class attendance has a significant impact on test mark. The coefficient on lecture attendance however was not significant, even at the 0.10 level, implying perhaps that lecture attendance does not have a significant impact in a multivariate framework. However, lecture attendance does appear to have a reasonably high correlation with class attendance, so the regression may be suffering from multicollinearity which has made the result not significant. However, multicollinearity must be occurring with another factor being ‘unhelpful’ for it to have a negative impact on the regression. The coefficient on revision lecture attendance was significant up to the 1% level, thus implying that while we can fairly sure that revision lectures have a significantly negative impact, there is scope for the fact that the null hypothesis is indeed correct (type I error) and that the result is not significant.

The F-test for the joint explanatory power of the independent variables yielded an F-statistic of 15.07. This is significant at the 0.01% level as it exceeds the critical value of 3.78. Hence we can reject the null hypothesis given in the appendix. This means that the explanatory variables have made a significant joint contribution to exam performance.

Question 3
To investigate whether there are differences in performance between the sub-sample of 2002 students and previous year’s students I have created 5 dummy variables and added them to the original equation, as shown by equations 6 and 2 in appendix 3. The first equation is as the restricted equation, as opposed to the unrestricted model in equation 2, because it imposes the F-test null hypothesis (see hypothesis 4, appendix 3 on equation 2). Hence in equation 2, the intercept is allowed to vary whereas it is not allowed to equation 1 and is assumed to be constant in all years.

Interpretation of coefficient

- The intercept in equation 3 can be interpreted as before, meaning that if you attended no lectures and had no A’s at A level you would score 56.97. This is slightly nonsensical in the sense that you would not have got onto the course if you did not score any A’s at A level.
- The coefficient of 0.14 on lecture attendance means that if you attended 1% more lectures you would get 0.14 out of 100 more in the exam centiles perilsus. The coefficient of 0.04 on A’s scored at A level means that if you get an extra A at A level you would get 0.04% more centiles perilsus. The dummy variables in this case have a slightly different interpretation. Basically they say how much the intercept will move up or down compared to the omitted category, the year 2000 students.
- The dummy variable coefficient on 1998 of -3.19 means that if you are a 1998 student, you will score a proportion of 1.19% less than if you were a 2002 student. The coefficient of -3.19 on the 2000 dummy variable means that you will score a proportion of 3.19% less than if you were a 2002 student.

Finally the coefficient of -0.85 on the 2001 dummy variable means that you will score a proportion of 0.85% less than if you were a 2002 student. These are shown in equations 5 to 6.
Comparisons of data across various groups

**Pure Economics degree vs. non pure Economics degree**: Students doing pure Economics degree scored 66.23 on average, while students doing a mixed-Economics degree scored 61.68 (very significant).

**Female vs. Male**: The average female students got 63.8, compared to 65.4 for male students.

**UK students vs. non-UK students**: On average, UK students gained 64.63 while non-UK students gained 66.38.

**Number of parents who attended university**: Those students whose parents never attended university achieved 64.12 on average, those with one parent attended university achieved 64.99 averagely, and those with both parents attended university achieved 65.59 on average.

To make some comments about these results, we need to break this up into sub-samples. Firstly we can break it up according to sex, as Siegfied and strand did. For males, table 2 shows that the mean score is 65.4%, which is higher than the corresponding score for females of 63.75%. This agrees with Siegried and Strand’s paper which claims males do better than females. However the standard deviation for males is lower than for females, 12.64% compared to 14.02%.

Use of lists by the Chinese student entails:

- Lower connector usage
- Fewer words needed to say the same thing
- Higher mean word length and lower mean sentence length
Interviews with lecturers

Importance of visuals

• Diagrams and formulae are ‘the spine of the essay’ (Economics)
• Including visuals helps students gain better marks as it avoids having to describe and introducing errors (Biology)
• The ‘challenge’ is ‘to marry the diagrams with the text’ (Economics)
• ‘there is no existing document out there which explains how to interpret their data’ (Biology)
• ‘the key writing skill for an economist is the ability to demonstrate in writing about a diagram an understanding of the analysis’ it represents (Economics)

Being concise

• Lecturers value writing which is ‘clear and concise’, and ‘succinct’ and dislike ‘verbosity’ (Engineering)
• ‘there’s never been a penalty for an essay that’s too short (Biology)
Conclusions

• Chinese students make greater use of visuals and lists than British students
• As all the assignments have achieved high scores, these differences are acceptable
• NB… EAP teachers do not always teach or encourage the use of visuals and lists – more training in disciplinary differences needed
• Need to read texts from the corpus to see what’s going on and what is missing/altered from the original texts
• Need disciplinary insights from lecturers (assignment-readers) and from students (assignment-writers)

Next steps…

• Establish ways of analyzing visual elements in assignments
• Talk to lecturers and students as to the range of acceptability in assignment layout/use of visuals/writing in lists
References