The Open University

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

ESRC project number
RES-000-23-0800

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

The corpora

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

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*=.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

Key in Chi-Economics	*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*
Key in Eng-Economics	*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*
Key in both	*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*

- 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

per 10,000 words	Chi Biol	Eng Biol	Chi Engin	Eng Engin	Chi Econ	Eng Econ
I	1	1	3	10****	9	14*
we	6	7	17	15	29	23
Total	7	8	20	25	38	37

* 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

	Chi-Economics	(raw)	Eng-Economics	(raw)
however	14	(55)	34****	(178)
therefore	17	(65)	23*	(122)
hence	8	(32)	16***	(86)
thus	12****	(44)	4	(20)
in contrast	2**	(7)	0	(1)

* 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

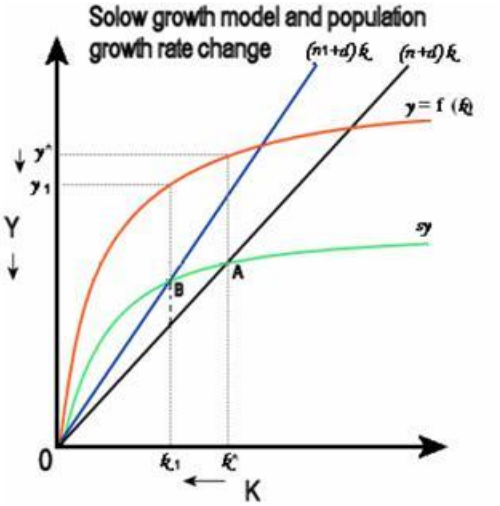
Keywords

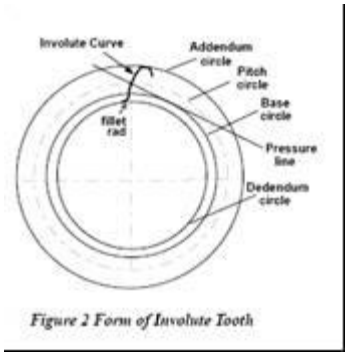
Key in Chi-Economics	*rate, model, output, formula, level, growth, curve, income, government, supply, students, population, dividends, per, reserves, consumption, T, aggregate, tax, Dutch, quantity, stock, portfolio, assets, inefficiency, competitive, capm, generation, repurchases, qtmark, asset, refer, cash, disposable, progress, deficit, income*
Key in Eng-Economics	*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*
Key in both	*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*

- 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…





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

Table.1

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).

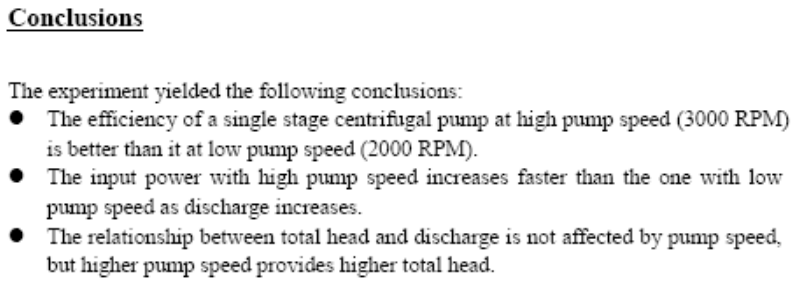
Visuals and lists

- Sales promotion:
  - Monthly promotions, according to customer sales and current interest
  - Discounts for bulk orders
  - Free P&P on orders over £25, encouraging bulk buying
  - Discounts for new businesses using us for the first time, on condition they use us for a minimum of two more orders
- Public relations:
  - User friendly website
  - Easy search tools within website, enabling you to find the exact bulb you want even for the non-engineering minded
  - Extensive "Help" and "FAQ" pages
  - 12hour guaranteed reply to email queries
  - Phone line for any further queries
- Selling
  - First class post as standard delivery
  - Next day delivery available on all orders for minimal extra
  - "Within six hours" delivery service available for a premium price
  - All transactions by Direct Debit over secure internet sites ensures fast turnaround and payment
  - Convenience and flexibility in not having to go to a shop, and always open

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.



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.

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).

Keywords relating to visuals and lists

L1& discipline	Chi-Biol	Chi-Bus	Chi-Econ	Chi-Engin	Chi-Food
Selected keywords	*#*		*growth*	*#*	*#*
	table		*curve*	*eq.*	*curve*
	data		*refer*	*according*	*referring*
	equation		*model*	*figure*	*statistical*
	figure		*per*		*deviation*
	graph		*output*		*numbers*

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

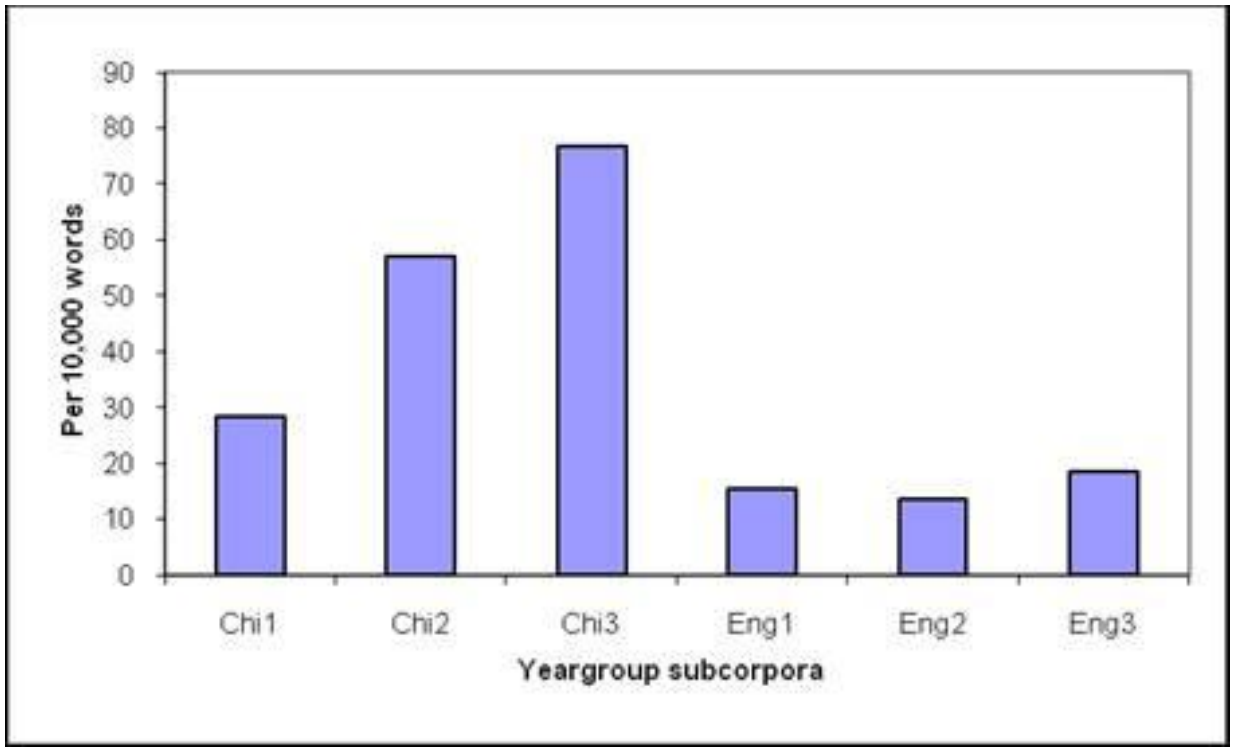
- **According to the** 3 **sets of data calculated above**… (6150d).

Visuals and lists

	Tables	Figures	Lists	Listlikes
Chi-Biol	15****	25****	1	4
Eng-Biol	5	13	2	6
Chi-Econ	1	14****	2*	25****
Eng-Econ	0	12	1	3
Chi-Bus	2	2	6*	129****
Eng-Bus	6**	6**	3	23
Chi-Food	20*	6	5	82****
Eng-Food	14	6	4	18
Chi-Engin	10*	21	7	53****
Eng-Engin	7	21	10	24

* p<.05
** p<.01
**** p<.0001

Listlikes by yeargroup



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

Text feature	Chinese, text 0434a	English, text 0067b
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 underling 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)

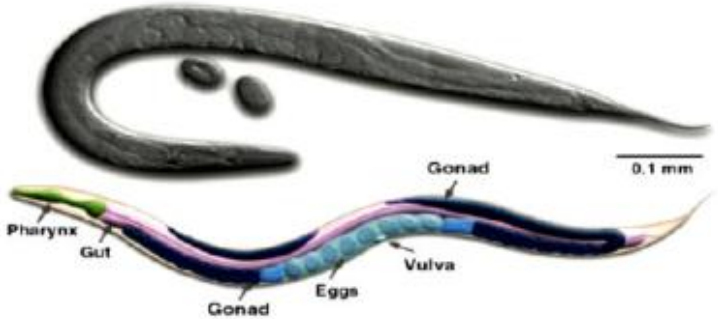


Figure 1 Adult *C. elegans* [1] Upper diagram: differential interference contrast image of an adult *C. elegans*. Lower diagram: anatomical structures of adult *C. elegans* (schematic drawing). Middle Left scale bar: 0.1 mm

The life cycle of *C. elegans* contains an embryonic stage, four larval stages (L1-L4) and an adult stage. (Figure 2) Molt (apolysis, 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

The role of maternal effect genes in the development of the nematode *Caenorhabditis elegans*

Maternally expressed genes are essential for the correct patterning and cell-fate determination in the early *Caenorhabditis elegans* embryo. The PAR proteins and MEX-5/6 are responsible for initial polarisation of the zygote, *skn-1* is required to specify the EMS fate, the bifunctional protein PIE-1 is required to maintain the totipotent germ cell lineage and specification of the AB lineage involves a system homologous to Notch in *Drosophila*. This review describes the current understanding of these molecular mechanisms in the specification of cell fates in the pregastrulation embryo.

Introduction

The potential of *Caenorhabditis elegans* as a model organism for the study of embryology emerged in the 1970s (Brenner, 1974). This free-living soil nematode is ideal for studying in the laboratory as it has a rapid period of embryogenesis (16 hours) and each worm has an invariant cell lineage, with exactly 959 somatic cells in the adult, which can be easily traced during development through the transparent cuticle (Sulston & Horvitz, 1977). *C.elegans* is a small round-worm, approximately 1 mm long, that lives for 2-3 weeks and can be fed on *Escherichia coli*, which allows large numbers to be conveniently raised in a Petri dish. The predominant adult form is hermaphroditic, containing both sperm and eggs and therefore reproduction is rapid, either by self-fertilization or by cross-fertilization with the rare males.

The genetics for *C.elegans* is advancing rapidly. It has a small genome at 8×10^7 bp and relatively few genes for a eukaryote – around 17,500. It was the first multicellular organism for which the genome was completely sequenced (*C.elegans* Sequencing Consortium 1999) and approximately 8,000 *C.elegans* proteins have already been matched to homologous human gene transcripts (Lai *at al.*, 2000). Specific mutants 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 hermaphrodite 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 ovoid is termed the anteroposterior (a/p) axis of the embryo.

Text 0434a, Chinese writer

Text 0067b, English writer

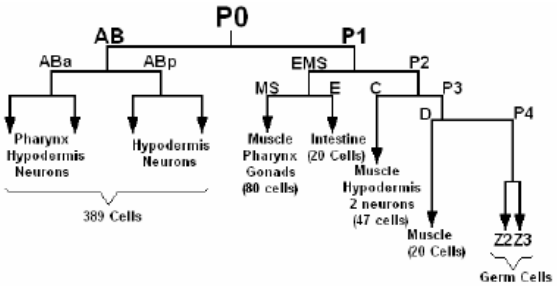


Figure 4 Formation of founder cells (modified from [2]). Asymmetrical division generate founder cells (AB, MS, E, C, and D) and germline lineage.

SETTING UP THE INITIAL ASYMMETRIES IN THE EMBRYO

Sperm entry - Initiation of Anterior-Posterior (A-P) polarity

During fertilization in *C. elegans*, the sperm not only triggers the production of the impermeable eggshell and the completion of oocyte meiosis I and II, but also play a crucial role in establishing the A-P polarity of the zygote (Figure 5). The sperm pronucleus moves to the nearest end of the oblong oocyte and defines it as the future posterior pole. [18] Pronucleus/centrosome complex (SPCC) of the sperm is crucial for initiating the A-P axis establishment by inducing polarized changes in both cytoplasm and cortex (during pseudocleavage).

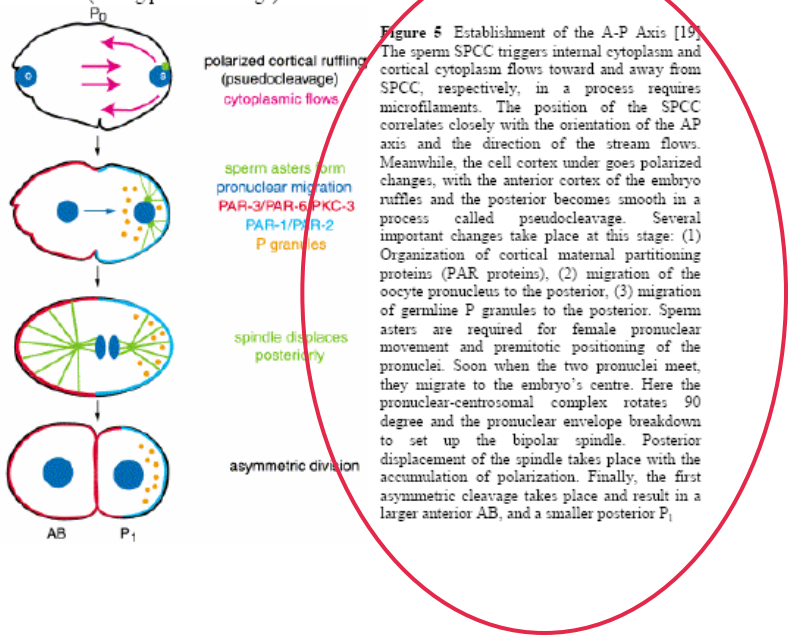


Figure 5 Establishment of the A-P Axis [19] The sperm SPCC triggers internal cytoplasm and cortical cytoplasm flows toward and away from SPCC, respectively, in a process requires microfilaments. The position of the SPCC correlates closely with the orientation of the AP axis and the direction of the stream flows. Meanwhile, the cell cortex under goes polarized changes, with the anterior cortex of the embryo ruffles and the posterior becomes smooth in a process called pseudocleavage. Several important changes take place at this stage: (1) Organization of cortical maternal partitioning proteins (PAR proteins), (2) migration of the oocyte pronucleus to the posterior, (3) migration of germline P granules to the posterior. Sperm asters are required for female pronuclear movement and premitotic positioning of the pronuclei. Soon when the two pronuclei meet, they migrate to the embryo's centre. Here the pronuclear-centrosomal complex rotates 90 degree and the pronuclear envelope breakdown to set up the bipolar spindle. Posterior displacement of the spindle takes place with the accumulation of polarization. Finally, the first asymmetric cleavage takes place and result in a larger anterior AB, and a smaller posterior P_1

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

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

Year 2002 vs. Year 2003 vs. Year 2004: The average mark for 2002, 2003 and 2004 students are 63.96, 61.70 and 69.19 respectively. Students in 2004 did better than Year 2002 and 2003. **These qualitative variables would have some impacts on the QTMARK, but whether they are statistically significant will be investigated later.**

Correlation matrices[2]

- It is found that QTMARK has strong POSITIVE relationships with variables ABILITY, ALEVELSA, ATTC, ATTL and strong **NEGATIVE** relationships with variables EXPALC and TOPB.
- The correlations between ATTC, ATTL and ATTR are very high. Therefore, **multi-collinearity** is an issue needed to think about later on.
- EXPALC has strong **negative** relationships with ATTC, ATTR and ATTL and strong **positive** relationships with TOPB. It could be explained that students who got drunk cannot 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 ALEVELS and ALEVELSA, while strong **negative** relationship with ABILITY and TOPB. This could be explained that the more TopB students went for, the fewer hours they spent on studying and the higher ability a student has the less 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

$$QTMARK=\alpha^1+\beta^1 ATTR+u^1$$

The following results are obtained after running the bivariate regressions in EViews[3]

$$QTMARK = 64.97023 - 0.002217ATTR$$

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.
- Slope coefficient of -0.0022 shows an inverse relationship between the attendance of revision lecture and exams 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 nearly omit its effect.
- **R-squared** value of about 0.000047 means that only 0.0047% of the variation in QTMARK 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^1

H0: $\beta^1 = 0$ (Proportion of revision lecture attended does not affect exam performance)

H1: $\beta^1 \neq 0$ (Proportion of revision lecture attended does affect exam performance)

Since the calculated t-value -0.13 is lower than the critical value of t-test at 5% significance level with 370 d.o.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

$$QTMARK= \alpha + \beta_1 ATTR + \beta_2 ABILITY + \beta_3 HRSQT + u$$

Modelling by OLS[4], we get:

QTMARK= 36.82527 + 0.105949ATTR+ 0.849900ABILITY+ 0.417658HRSQT

Interpretation of the regression results:

[2] Refer to "Correlation matrices for those quantitative variables" in the Appendix.
[3] Refer to the Appendix for the Bivariate regression results tables.
[4] Refer to the Appendix to find out the Multivariate regression results table.

Text 0155a, Chinese writer

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 by 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, revision or standard lectures, you would score 49.33%

Tests (shown in appendix 2)

The coefficient on class attendance was significant at the 0.01% level implying that in the multiple regression models, class attendance has a significant impact on test mark. The coefficient on lecture attendance however was not significant, even at the 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[3], 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 be 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 13.07. 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 intercept dummy variables and added them to the original equation, as shown by equations 1 and 2 in appendix 3. The first equation is known 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 coefficients

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 ceteris peribus. 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 ceteris peribus. 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 1999 of -1.19 means that if you are a 1999 student, you will score a proportion of 1.19% less than if you are 2002 student. The coefficient of -5.19 on the 2000 dummy variable means that you will score a proportion of 5.19% less than if you were a 2002 student. Finally the coefficient of -6.85 on the 2001 dummy variable means that you will score a proportion of 6.85% less than if you were a 2002 student. These are shown in equations 3 to 6.

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

Text 0202j, English writer

Extracts from Economics assignments

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

- British Council. (2010). China Market Introduction. Retrieved 17/08/2010, from http://www.britishcouncil.org/eumd-information-background-china.htm
- Heuboeck, A., Holmes, J. & Nesi, H. 2007 *The Bawe Corpus Manual.* Retrieved from http://www.coventry.ac.uk/researchnet/d/505/a/5160.
- Hewings, A. (2004). Developing discipline-specific writing: An analysis of undergraduate geography essays. In L. J. Ravelli & R. A. Ellis (Eds.), *Analysing Academic Writing: Contextual Frameworks* (pp. 131-152). London: Continuum.
- Hyland, K. (2008). As can be seen: Lexical bundles and disciplinary variation. *English for Specific Purposes, 27*(1), 4-21.
- Kress, G., & Van Leeuwen, T. (2001). *Multimodal Discourse*. London: Arnold.
- Leedham, M. (2009) 'From traditional essay to 'Ready Steady Cook' presentation: reasons for innovative changes in assignments' In Active Learning in HE.
- Nesi, H. and S. Gardner (2006) 'Variation in Disciplinary Culture: University Tutors' Views on Assessed Writing Tasks', in Richard Kiely, Gerald Clibbon, Pauline Rea-Dickins, and Helen Woodfield, (eds) *Language, Culture and Identity in Applied Linguistics* (British Studies in Applied Linguistics, 21) London: Equinox Publishing.
- Scott, M. 2008. WordSmith Tools v.5. Oxford University Press.