AC 2011-1565: FREQUENCY ANALYSIS OF TERMINOLOGY ON ENGINEERING EXAMINATIONS

Chirag Variawa, University of Toronto

Chirag Variawa is a Ph.D. candidate in the Department of Mechanical and Industrial Engineering at the University of Toronto. His research interests include maximizing inclusivity, accessibility and usability of engineering education via universal instructional design and innovative instructional methods. He is an active Canadian member of the SCC division of ASEE, co-chair of the Leaders of Tomorrow (Graduate) program and teaching assistant in the Faculty of Applied Science and Engineering. He received his B.A.Sc. (2009) from the Department of Materials Science and Engineering, University of Toronto.

Susan McCahan, University of Toronto

Susan McCahan is a Professor in the Department of Mechanical and Industrial Engineering at the University of Toronto. In addition, she is currently the Chair of First Year for the Faculty of Applied Science and Engineering. She received her B.Sc. from Cornell University (1985), and M.S. (1989) and Ph.D (1992) degrees from Rensselaer Polytechnic Institute in Mechanical Engineering.

Frequency Analysis of Terminology on Engineering Examinations

Introduction

The engineering student population is becoming increasingly diverse in recent years.[1,2] As a result, the diversity of background experience and vocabulary that students bring with them to university is increasing as well. As these students integrate within engineering institutions, they may face issues of inclusivity and accessibility to course material because of their diverse backgrounds. One such dimension that particularly impacts student inclusivity is that of language.[3] Students may face barriers to learning when the language of instruction and assessment does not accommodate differences in learner characteristics. The problem is that students may actually have a different corpus of language than instructors assume they have. For example, when a student encounters a term that is unfamiliar to them, the word creates a barrier to understanding. This barrier may inhibit learning, or compromise the validity of assessment if the student's lack of understanding is not addressed. Some vocabulary (course specific vocabulary) is explicitly taught. However, when unfamiliar vocabulary is used, and not explicitly taught, it creates a misalignment between the learning environment and the learner. The learner experiences this as a barrier to accessibility of the learning environment

A potential solution to the issue of inaccessible language might appear to be the use of plain language. Plain language is the notion that clear and simple language is the most accessible and logical way of communicating with one another. There is plenty of literature in this area, and there are several studies that show the benefits of using plain language.[4-6] However as educators, we want our students to develop a deep and robust vocabulary as part of their engineering education. This is particularly important because the mastery of technical and professional corpora of language is beneficial for students and practicing engineers alike. As a result, educators cannot simply use plain language at an elementary level to address this issue, but instead need to investigate the issue of inaccessible language in their curricula.

Our hypothesis is that word familiarity is correlated with word frequency. If this is true then words that appear frequently in teaching materials are better understood by students, and words that appear infrequently are more likely to be unfamiliar. The first step in this investigation is analyzing the frequency of words in a typical engineering classroom. Specifically, we believe that this approach will provide some insight on the issue of inaccessible vocabulary used in engineering education and also begin to characterize the nature of the language corpus used in engineering education. This study measures the frequency of words in one particular type of learning material, undergraduate final exams, because this method of closely-supervised assessment is common in engineering education and provides a substantial database of language to analyze.

Some analysis techniques in the area of vocabulary frequency-analysis were developed by C.J. van Rijsbergen.[7,8] Primarily, his work comments on the use of Zipf's Law to understand the statistical distribution of words in language. Zipf's law states that the most frequent word in an article of text will appear twice as frequently as the second most frequent word, and four times as frequently as the third most frequent word, etc.[9] Thus, the expected result of a frequency analysis is a hyperbolic curve with a narrow range of frequently-appearing words and a broad range of infrequently occurring words. To better understand the validity of this theorem, Li performed a study using a uniform distribution of all 26 letters, plus a "space" character to study the effect of Zipf's law in different cases. His approximation established that the law is indeed valid no matter what vocabulary is used, but that its effect is more pronounced in natural language.[10] Natural language is a term used in the literature that describes vocabulary that has evolved in an unpremeditated fashion. The rough theory behind this phenomenon is that humans often mix frequent and infrequent words that may or may not have meaning individually.[11,12] Additionally, this theory also helps to explain why the word "the" (and "vowel-less" words) are generally excluded from frequency analysis studies; the word "the" is the most common word in the English language.[7,13] Overall, Zipf's law is one concept we can use to interpret the data acquired from the frequency analysis of words.

Methodology

The objective of the current work is to develop lists of words ranked by frequency for a set of engineering course final exams. These lists will then be processed in two ways. First, the word lists will be inputted to a database program so that the frequency and rank can be accurately matched to its corresponding word and exam. Second, the lists will be plotted graphically to determine overall trends in the data. The expected output from this process will be a dataset of vocabulary with each word being tagged by rank and frequency.

The study investigates the frequency of words used on engineering examinations at the University of Toronto. Final examinations were chosen for this study for several reasons. The database of final exams is readily available. At this institution final exams from previous years are posted on a publicly-accessible website so that students can use them as study aids. Also, students are not able to access assistance during an exam, which means that they must rely on their a priori vocabulary to make sense of the questions. And as a critical assessment in a course, the exam should be testing the student's understanding of the course concepts rather than the student's non-discipline-specific vocabulary. Presumably, the instructor has taken this into account when developing the exam. Finally, every exam in this program is the same duration, 2.5 hours, which allows for some common basis of comparison (e.g. number of words on the exam).

These exams are posted in PDF format, and include information about the course and instructor. We started by downloading the most recent exams from the freshman courses in Materials Science Engineering (MSE). An advantage of using this set of exams is that these courses are the same, or very similar to, courses taken by other freshman engineers at many institutions. Further, the authors have experience with the content and assessment objectives of each exam, making is easy to identify vocabulary that is explicitly taught in each course. Additionally, using electronic exams enabled a computational solution to performing the frequency analysis.

To perform the frequency-analysis, several software tools are used. Each exam was first processed using Adobe® Acrobat Pro v.9.4.1 to make the text searchable electronically. The optical character recognition (OCR) engine in this software is responsible for converting static images into text. The main advantages of using this software are that it dramatically reduces the amount of time and effort required to input text for the frequency computation, and reduces human error in data entry. A disadvantage of this approach, however, is that each word is not vetted by a human prior to entry. This means that typos in the original document are treated as actual words, and these eventually become part of the compiled database. Additionally, there may be cases where the software disregards disfigured words because they are unintelligible to process. In the case of distorted words, the authors made corrections manually.

The next step in this process was to use a program called Hermetic Word Frequency Counter Advance v.12.45. The program calculates the frequency of each word, and outputs the data as a text file which includes rank, word, and word frequency, and a unique identifier for each word. During this process, we instructed the program to disregard particular words, i.e. exclude particular words from the computation. Specifically, the program ignores specified character strings that are illogical and may confound the results (i.e. words that have less than 2 characters, contain a hyphen, that are just repetitions of the same letter, or which lack a vowel or 'y'). One reason for removing these words is to reduce the likelihood of an erroneous word, such as equations and variable names, being processed, and to reduce the clutter in the database. One disadvantage of this software system, however, is that words with prefixes and/or suffixes are considered to be their own unique word. For example, the word "gear" would be considered different from "gears". This limitation is one that the literature recommends be addressed, but there is currently no automated method for accomplishing this, and no clear systematic approach described in the literature. So we have not manipulated the data to combine the results for variations of the same word. Additionally, we also acknowledge that the exam sample size (N=9) is presently small. However, the word lists are substantive (n=565, in total). This provides a large vocabulary sample that is indicative of the language corpus used in introductory engineering courses.

The word lists that were produced from this process are sorted by frequency. The data was then plotted to produce graphs comparing the vocabulary frequency, and this was used as the basis for comparing different exams to one another. Further, the frequency distributions for each exam were examined statistically to understand how different exams compare and how they compare to natural language. This method produces data that can be mined in a variety of ways to better understand the language we use in the engineering learning environment.

Results and Discussion

The frequency analysis produced nine datasets containing ranked frequency distributions of the words used on each exam. The nine course exams analyzed came from:

- Calculus I
- Calculus II
- Linear Algebra
- Mechanics
- Introduction to Materials Science
- Physical Chemistry
- Fundamentals of Computer Programming
- Electrical Fundamentals
- Engineering Strategies and Practice (ESP)
  which is an introductory design and communications course

Our preliminary data shows that these distributions can be roughly placed along a spectrum, with "Calculus I" (Mathematics) representing one extreme and "Engineering Strategies and Practice" (ESP) representing the other, as seen in Figures 1 and 2, respectively. The word-frequency distributions from the other courses fall between these two extremes. Samples from the raw word sets for these two exams are included in the appendix as Table 1 and 2, respectively.

The data shows that the occurrence of words which we might assume are very familiar, such as "name", "clear", or "length", are not particularly frequent (nor consistently infrequent). Additionally, the data illustrates that mathematics exams, especially Calculus I, generally have fewer words than other exam types. The data also demonstrates that all exams have a roughly hyperbolic distribution of words per Zipf's law; some words occurring extremely frequently, and most occurring only once.

Figures 1 and 2 show the word frequency distribution of Calculus I and ESP, respectively (excluding the word "the"). The vertical axis is the "occurrence percentage". This normalized value is calculated by dividing the number of occurrences of a specific word by the total number of words on that exam. This number shows how common the word is to the particular exam. The horizontal axis is the rank of any unique word, as referenced in Table 1 and 2 in the appendix. Basic statistical analysis was also carried out for the entire dataset for all nine exams. The minimum number of words used is 91 (Calculus I), and the maximum is 525 (ESP). The mean and standard deviations for this dataset are 282.7 and 164.3, respectively. This indicates that there is a large variability in the word count for the exams studied; some exams have a much higher count than others. Overall, the data offers a preliminary look at the way vocabulary is utilized in engineering learning materials.

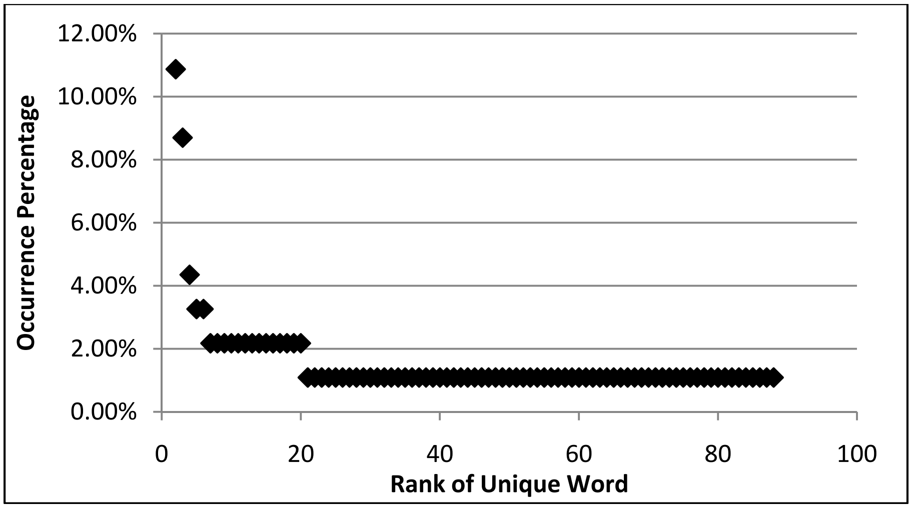

Figure 1. Shows the word frequency distribution for the Calculus I final exam (Mathematics). The total number of unique words is 91.

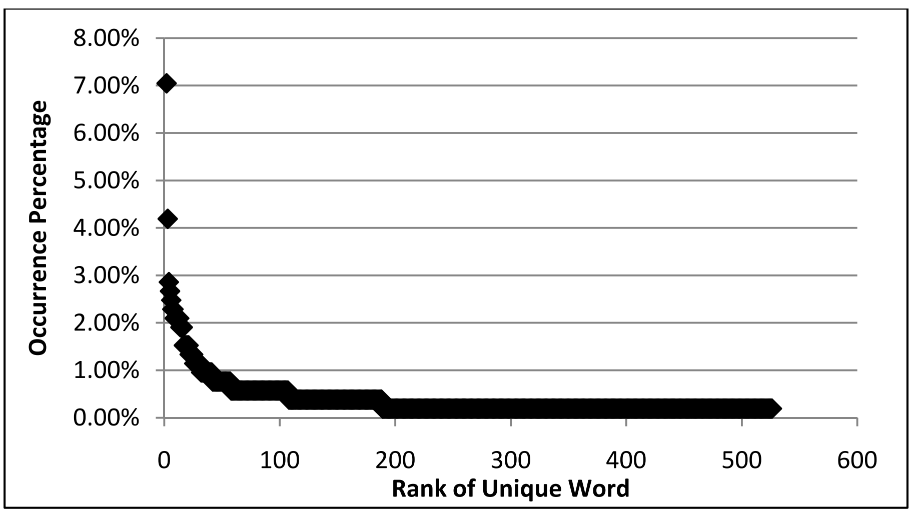

Figure 2. Shows the word frequency distribution for the Engineering Strategies and Practice final exam. The total number of unique words is 525.

The data show that the correlation between the independent and dependent axes is not linear, it is hyperbolic. This follows Zipf's law, demonstrating that a small number of words are used much more frequently than all others. The frequency distribution also reveals that exams that have fewer words have a weaker hyperbolic correlation between occurrence percentage and rank. According to Zipf's law, a weak correlation means that the language used in these exams has low similarity to natural language (NL). Zipf's law states that the frequency of any word is proportional to its rank in the frequency table.[9] Our data implies that the mathematics exams, especially Calculus I in our sample, use language that departs from natural language. A result which is similar to Li's study[10] mentioned earlier,. Examining the word lists from the math courses and design course supports this. Words that are common in natural language, such as "then", "if", and "but" appear infrequently on the math exams. Words that appear frequently on the math exams include "point", "sum", and "choice" which are less usual in natural language.

The words contained in the Calculus I data set also have fewer prepositions than ESP, further suggesting that it contains less NL-based questions. Generally, NL contains a high number of prepositions used to accurately characterize the word (generally noun) succeeding it.[14] As a result, the quantity of prepositions can be used in combination with frequency distributions, raw word sets, word counts and occurrence-percentages to predict whether an exam contains NL-based questions. However, when we analyse these findings together, we get an even better sense of whether the examination uses NL that we assume all students would know.

The results of this study can be situated in the context of the existing literature. Luhn's work in the field of information retrieval suggests methods of data mining that can be applied to word frequency datasets.[7,15, 16] Luhn worked extensively with information retrieval technologies, and suggests ways of data mining for accurate retrieval based on input queries. Specifically, he suggests that words be assigned tags and weightings. Tagging, for example, can be used to distinguish unique word definitions, e.g. allow "Apple®" the company name to be distinguished from "apple" the fruit. Further, words that are variations of the same root word can be assigned the same tag. This reduces the clutter in the dataset because synonymous words are deleted, assuming the tagging has been done carefully. Weighting, however, means that we assign values to words; the values could be assigned based on a particular set of criteria. For example, we could assign terms that are less familiar to our students a higher weighting than terms that are more familiar, if there was data on familiarity. In general, tagging and assigning weights are both examples of grouping techniques that help condense the dataset into more manageable units. Used together, these methods may help to identify words that combine frequent use with low familiarity, i.e. words that may pose the most frequent, and significant learning barriers for students. This is an important consideration if we want to distinguish inaccessible terms from accessible ones. However, this separation first requires that we understand the characteristics of the learner's vocabulary a priori. Knowing these characteristics, it is then possible to use the frequency distribution graphs that have been developed to isolate regions where inaccessible terms are most likely to appear. The literature suggests that upper and lower cut-off points can be defined on a word frequency graphs.

However, it is not clear how to best apply this methodology if the learner's a priori vocabulary is both unknown and continuously shifting. It may be possible with the growing usage of electronic textbooks for students to identify unfamiliar words as they are studying a subject and have this data collected automatically. We can imagine a system that works much like a spell checker to identify potentially unfamiliar and problematic words based on word frequency in a particular discipline or class. However, to understand the current data in the framework of learner characteristics it is necessary to establish a proxy system that makes use of some other feature that is common to inaccessible vocabulary in order to bring it to the attention of the instructor and the student. Further, it is important to understand the limitations of these approaches so that we can more-fully elucidate the issues with word frequency analysis.

Van Rijsbergen provides a critique that organizes several approaches into a framework that can be used to understand word frequency analysis better and suggests, at least minimally, how to begin to develop a proxy system.[7] Specifically, his critique is important because it articulates the limitations of this work while informing a potential direction for the analysis. Van Rijsbergen explains how prefixes and suffixes affect the meaning of words. Moreover, an understanding of this issue allows us to remove related words to simplify the resulting dataset. For example, the removal of “ual” from “factual” retains the meaning in the root, but this is not true if “ual” is removed from “equal”.

In Van Rijsbergen’s interpretation of Luhn’s work, he establishes that most unique intermediate terms appear between the upper and lower cut-off points, as seen in figure 3. In our results, this range includes words such as “coexistence”, “conversion” and “dilemma”. These words do, from a purely subjective perspective, appear to be potentially more unfamiliar or less accessible for students. Moreover, these words may be more challenging than words such as “marks” or “thanks” which are very frequent or very infrequent. That is to say, although this is a blunt approach that may capture some very familiar terms, or leave out some unfamiliar terms, there appears to be some promise that inaccessible language can be bounded, to some degree, by using this word-frequency analysis technique. However, more work and a larger sample size will be required before a definitive conclusion is possible. It is also not clear yet where exactly to draw the cut-off lines.

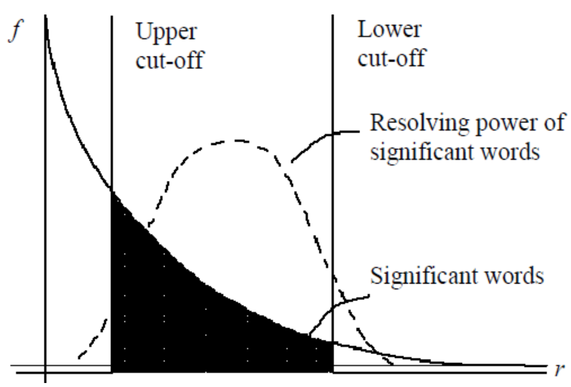

Figure 3. Shows how significant terms are likely located between the upper and lower cut-off regions. (Reproduced here[7,16])

At present, the accuracy of finding unfamiliar and inaccessible language is low. It is difficult to predict where these inaccessible terms are simply from browsing and comparing the graphs in Figures 1 and 2 alone. Our initial hypothesis was that inaccessible terminology would be used less frequently than accessible terminology. However, the individual data sets do not support this hypothesis. We found that unfamiliar and inaccessible terms are not necessarily infrequent. Rather, less accessible words may occupy a region that is intermediate between high frequency and low frequency. Further, the characteristics of unfamiliar language are vague; it is difficult to predict where these terms are without further and more in-depth work.

Such work could involve compiling a larger set of exams, reducing cluttering of the data by removing pre-/suffixes, using tagging as suggested by Luhn's work and comparing individual exams to an amalgamated dataset of all exams. At present, this small dataset is useful for exploratory work in a specific area. However, having a larger dataset can help to better assess the hypothesis. In addition, reducing the clutter in the database by focussing on the root word, rather than the form that includes pre-/suffixes, can assist in compacting the dataset. This is particularly useful in maintaining the integrity of the database because having multiple permutations of the same word still retains the same basic meaning, but adds to the overall word count of the exam. Also, comparing individual or groups of exams to an amalgamated dataset of all exams or existing natural language datasets might yield interesting results. This comparison may identify how a given exam or group of exams (for example, design courses) compares to the general characteristics of vocabulary used in these materials. Discussing the common features of these exams versus the large dataset may yield information about how a specific type of course might be more or less likely to have unfamiliar, inaccessible language for an undergraduate engineering learner population.

This is a first exploratory step in a line of study that informs an approach that might make engineering education more accessible for the majority of students. As such it is situated in a Universal Instructional Design (UID) approach to improving the learning environment for students.[17] However, it should be noted that there will be limitations to any set of results or remediation strategy that is developed from this work. First, there remains a portion of learners that are "high-risk". This population includes learners who require specialized individual attention or accommodation. For example, simply making vocabulary more familiar will not remove the need for accommodation for students with learning disabilities, but it may make the learning environment somewhat more accessible for these students. Another limitation to the applicability of this research is that vocabulary is not the only barrier to accessibility in the engineering classroom. There are many dimensions to learner characteristics that impact accessibility.[3]

There are, however, a number of advantages to finding and mitigating inaccessible vocabulary. Using accessible language may assist students who would not otherwise self-identify as people who face barriers in the learning environment. This is related to the "curb cut" effect mentioned frequently in the UID literature.[17] Overall, making language more accessible helps a diverse learning population feel more included in an environment conducive to professional skills development. This logic is often used in the Universal Instructional Design (UID) literature. UID describes principles that make the learning environment more accessible to students. For example, encouraging clarity and flexibility in the delivery of instructional material has a positive effect on a variety of students each having different learning ability characteristics.

Inclusivity can also potentially encourage greater student involvement in the learning process. Language is often cited as an issue in the literature on inclusivity.[3] Further, understanding language supports and encourages the development of a robust professional vocabulary while maintaining the integrity of the course learning objectives.

Conclusions

Language can be one dimension of some inclusivity and accessibility issues students' face in engineering education. Identifying vocabulary that might be unfamiliar and inaccessible has many benefits for all students. It helps students overcome learning barriers, while giving instructor's information they can use to help students develop a robust professional vocabulary.

Frequency analysis of language has several limitations, but this exploratory study has shown some interesting results. Specifically, the word frequency and word sets should be used jointly to determine the extent to which natural language-based questions are present on a given exam. In addition, the preliminary data show that potentially more arduous words may occupy a region intermediate between high and low-frequency. However, it is clear that work in this area is promising, and may inform an approach that identifies inaccessible words using frequency analysis after more rigorous studies. At present though, we need to determine specific criteria to help focus our search for potentially inaccessible vocabulary.

The applicability of accessible language in engineering pedagogy is profound. Using a UID approach, we can create more inclusive learning environments that are more flexible and can accommodate different learner characteristics. Our future work will investigate ways of improving the process of finding and mitigating inaccessible language used in all levels of engineering education, in addition to making the environment more accessible and inclusive for students.

References

[1] Synergies (2008 Annual Report). Rep. National Action Council for Minorities in Engineering. Web. <http://www.nacme.org/user/docs/NACME_AnnualReport2008.pdf>.

[2] Vision, the NACME Continuum (2010 Annual Report). Rep. National Action Council for Minorities in Engineering. Web. <http://www.nacme.org/user/docs/NACME_AR_2010_FINAL.pdf>.

[3] Variawa, C., and S. McCahan. 2010. Design of the learning environment for inclusivity. Proceedings of the 2010 American Society for Engineering Education Annual Conference and Exposition. Louisville, KY.

[4] Bello, D. President Signs 'Plain Language' Bill into Law. Safety Health. 182.6 (2010): 21.

[5] Petelin, R. Considering Plain Language: Issues and Initiatives. Corporate Communications. 15.2 (2010): 205-16.

[6] Harper, R., and D. Zimmerman. 2009. Exploring Plain Language Guidelines. Proceedings of the IEEE International Professional Communication Conference. Honolulu: HI.

[7] Van Rijsbergen, C.J. 1979. Chapter 2: Automatic Text Analysis. Information Retrieval. 2nd Edition, Butterworth: London. pp. 10-15. Web. <http://www.dcs.gla.ac.uk/Keith/pdf/Chapter2.pdf>.

[8] Lease, M. 2007. Natural Language Processing for Information Retrieval: The Time is Ripe (again). Association for Computing Machinery, Inc. New York:NY.

[9] Zipf, H.P. 1949. Human Behaviour and the Principle of Least Effort. Addison-Wesley, Cambridge:MA.

[10] Li, W. Random Texts Exhibit Zipf's-Law-Like Word Frequency Distribution. IEEE Transactions on Information Theory. 38.6 (1992): 1842-5.

[11] Bloom, L. Cognition and the Development of Language. Language. 50.2 (1974): 398-412.

[12] Saffran, Jenny R., et al. Incidental Language Learning: Listening (and Learning) Out of the Corner of Your Ear. Psychological Science. 8.2 (1997): 101-5.

[13] The Linguistics Encyclopedia. Ed. Kirsten Malmkjær. 2nd ed. 2002. Routledge, New York.

[14] González B., J., Pazos R., Gelbukh, A., Sidorov, G., Frairer, H., and I. Cruz. 2007. Prepositions and Conjunctions in a Natural Language Interfaces to Databases. Springer, Berlin, pp. 173-182.

[15] Luhn, H.P. The automatic creation of literature abstracts. IBM Journal of Research and Development. 2.0 (1958): 159-165.

[16] Schultz, C.K. 1968. H.P. Luhn: Pioneer of Information Science - Selected Works. Macmillan, London.

[17] Bowe, F. 2000. Universal design in education: Teaching nontraditional students. Bergin & Garvey, Westport, CT.

Appendix

This section includes all of the word sets for the two exams analyzed in this paper, Calculus I and ESP. The left column shows rank, the center column shows the word, and the right column indicates the frequency. Note: "the", which occurred 17 times for Calculus I and 121 times for ESP, is excluded from both tables.

Table 1. Word set for Calculus I.

Rank	**Word**	**Freq**
1	page	10
2	find	8
3	and	4
4	marks	3
5	using	3
6	area	2
7	between	2
8	derivative	2
9	for	2
10	function	2
11	let	2
12	lim	2
13	name	2
14	ofl	2
15	solid	2
16	that	2
17	total	2
18	volume	2
19	x-axis	2
20	allowed	1
21	applied	1
22	april	1
23	areas	1
24	around	1
25	arqund	1
26	calculators	1
27	calculus	1
28	check	1
29	circular	1
30	closest	1
31	cohen	1
32	cos	1
33	created	1
34	curve	1
35	cylinder	1
36	defined	1
37	definition	1
38	dimensions	1
39	duration	1
40	engineering	1
41	evaluate	1
42	evaluatef	1
43	exam	1
44	examiner	1
45	exists	1
46	faculty	1
47	family	1
48	final	1
49	fits	1
50	from	1
51	generated	1
52	given	1
53	graph	1
54	here	1
55	hours	1
56	integrals	1
57	largest	1
58	limit	1
59	limits	1
60	line	1
61	markers	1
62	mat	1
63	minutes	1
64	monday	1
65	not	1
66	number	1
67	only	1
68	please	1
69	point	1
70	question	1
71	radius	1
72	riemann	1
73	right	1
74	rotated	1
75	rotating	1
76	science	1
77	show	1
78	sign	1
79	sin	1
80	sphere	1
81	standard	1
82	student	1
83	sums	1
84	surface	1
85	then	1
86	this	1
87	toronto	1
88	university	1
89	when	1
90	whether	1
91	x-hlo	1

Table 1. Word set for Engineering Strategies and Practice

Rank	Word	Freq
1	and	37
2	for	22
3	number	15
4	you	14
5	this	13
6	page	12
7	what	12
8	are	11
9	name	11
10	student	11
11	that	11
12	your	11
13	design	10
14	use	10
15	why	10
16	engineering	8
17	explain	8
18	from	8
19	marks	8
20	type	8
21	exam	7
22	law	7
23	not	7
24	write	7
25	each	6
26	other	6
27	total	6
28	will	6
29	with	6
30	would	6
31	all	5
32	alternative	5
33	decision	5
34	ethics	5
35	his	5
36	now	5
37	pedal	5
38	two	5
39	using	5
40	word	5
41	can	4
42	describe	4
43	devices	4
44	equation	4
45	how	4
46	improve	4
47	lemessurier	4
48	matrix	4
49	one	4
50	profit	4
51	safety	4
52	section	4
53	sentence	4
54	want	4
55	which	4
56	years	4
57	answer	3
58	answers	3
59	any	3
60	below	3
61	between	3
62	bin	3
63	company	3
64	correct	3
65	correction	3
66	costs	3
67	current	3
68	driving	3
69	example	3
70	first	3
71	flaps	3
72	following	3
73	give	3
74	human-tech	3
75	identify	3
76	its	3
77	journal	3
78	kong	3
79	ladder	3
80	large	3
81	level	3
82	list	3
83	make	3
84	may	3
85	model	3
86	need	3
87	objectives	3
88	only	3
89	open	3
90	please	3
91	problem	3
92	received	3
93	required	3
94	road	3
95	service	3
96	should	3
97	space	3
98	sure	3
99	team	3
100	terms	3
101	these	3
102	three	3
103	toronto	3
104	university	3
105	while	3
106	widgets	3
107	actions	2
108	agency	2
109	allowed	2
110	analysis	2
111	axes	2
112	back	2
113	based	2
114	blank	2
115	booklet	2
116	case	2

117	city	2
118	client	2
119	comment	2
120	communications	2
121	constraints	2
122	contract	2
123	cycle	2
124	definition	2
125	development	2
126	did	2
127	different	2
128	dilemma	2
129	does	2
130	dogs	2
131	end	2
132	eng	2
133	environment	2
134	environments	2
135	faculty	2
136	final	2
137	flaws	2
138	foot	2
139	functions	2
140	garbage	2
141	has	2
142	however	2
143	include	2
144	includes	2
145	indicate	2
146	information	2
147	label	2
148	liability	2
149	life	2
150	line	2
151	maximize	2
152	mechanism	2
153	must	2
154	new	2
155	objective	2
156	ontario	2
157	operating	2
158	out	2
159	part	2
160	phone	2
161	process	2
162	provided	2
163	quadrants	2
164	questions	2
165	re-design	2
166	read	2
167	requirements	2
168	same	2
169	second	2
170	sentences	2
171	spend	2
172	stages	2
173	statement	2
174	steps	2
175	study	2
176	term	2
177	them	2
178	third	2
179	time	2
180	title	2
181	toys	2
182	ultimate	2
183	users	2
184	value	2
185	was	2
186	who	2
187	above	1
188	accessibility	1
189	accommodate	1
190	accredited	1
191	accurate	1
192	acid	1
193	added	1
194	additional	1
195	administered	1
196	advertised	1
197	agree	1
198	aids	1
199	air	1
200	aisle	1
	…	
210	applicable	1
220	assumptions	1
230	better	1
240	calculator	1
250	cent	1
260	closed	1
270	conditions	1
280	currently	1
290	degree	1
300	discarded	1
310	effect	1
320	entry	1
330	facultative	1
340	four	1
350	having	1
360	impact	1
370	ivey	1
380	lines	1
390	mean	1
400	needing	1
410	our	1
420	piloted	1
430	production	1
440	rating	1
450	relevant	1
460	sehtences	1
470	slots	1
480	still	1
490	than	1
500	units	1
510	ways	1
520	worked	1
	…	
525	year	1