When BAWE meets WELT:
The use of a corpus of student writing to develop items for a proficiency test in grammar and English usage

Gerard Paul Sharpling

University of Warwick | United Kingdom

Abstract: This article reports on the use of the British Academic Written English (BAWE) corpus as a source for developing test items for the Grammar and English Usage section of the Warwick English Language (WELT) test in 2007. A key feature of this newly designed multiple choice grammar test was its use of student-generated writing. The extracts used for the re-designed test were derived directly from the BAWE corpus, as opposed to text books, published sources or indeed, simulated extracts of academic writing devised by test developers, which had been the case previously. The rationale for using the BAWE corpus for language test design is outlined, with a particular focus on the attributes of the students' writing within the corpus, and the inclusion of both first and second language writing. The challenges involved in developing grammar test items based on BAWE corpus data are also presented. While the procedures set out in the paper were undertaken within a specifically British higher education setting, it is hoped that the research will be of interest to test developers and/or researchers in writing skills in other academic settings worldwide.

Keywords: corpus, BAWE, WELT, academic writing, language test design, English for Academic Purposes, L1/L2 writing

1. Introduction

This paper reports on the use of the British Academic Written English (BAWE) corpus as a source for designing and developing a test of grammatical proficiency. BAWE is a searchable collection of written assignments, first set up in 2001. The pilot corpus was completed in 2004 (see Nesi, Sharpling & Ganobcsik-Williams, 2004) and contains about one million words of text, in the form of 500 student assignments ranging from 1,000 to 5,000 words, representing twenty departments at Warwick University. Additional funding to expand and analyse the BAWE corpus was obtained between December 2004 and December 2007 for a project entitled "An investigation of genres of assessed writing in British Higher Education". Texts for the BAWE corpus were supplied by students from four universities within the UK. This led to the formation of a new BAWE corpus, consisting of approximately 3000 assignments. The corpus contains assignments taken from four disciplinary areas: Arts and Humanities, Social Sciences, Physical Sciences, and Life and Medical Sciences (Alsop & Nesi, 2009). Just over 800 of the assignments (30%) were written by L2 writers (Hyland, 2008). Tutors from contributing departments were also interviewed, in order to find out more about departmental practice (Nesi & Gardner, 2006). Assignments in the corpus were arranged according to written genres, based on tutors' perceptions, evidence from departmental documents, the wording of assignment titles and specific linguistic features associated with the assignments. The corpus has recently been made available to researchers through the Oxford Text Archive.[1]

The Warwick English Language Test (WELT) is an English language proficiency test for candidates across the world applying to be accepted for further study by Higher Education (HE) institutions within the United Kingdom. The test was first introduced in 1989, and until recently, was widely accepted as an alternative to well-known international proficiency tests such as IELTS and TOEFL. WELT has in recent years been recognised by over 30 universities, and its candidature has averaged 2000 per year, from thirty different countries. The test consists of three papers: a one hour, 100-item multiple choice (MCT) Grammar and English Usage paper; a 45-minute writing paper; and a one hour, thirty item multiple choice reading test. A letter grade (A to E) for each paper is attributed, to indicate whether a candidate has reached the required standard of language proficiency to proceed to their chosen academic programme.[2] WELT is untypical of other proficiency tests, insofar as it has a separate test of grammatical competence.[3] The first half of the test (questions 1 to 50) assesses candidates' ability to recognize common errors and choose appropriate academic language in context. Candidates are also required to distinguish grammatical from ungrammatical sentences. In the second part of the test, (questions 51 to 100) candidates are required to recognize how best to complete sentences or short paragraphs from four given choices. The grammatical areas covered by the test are similar to the schema outlined by Purpura (2004), which sees grammatical ability as being subdivided into "grammatical

form", "grammatical meaning" and "pragmatic meaning." This tripartite view of grammar is presented more fully in Table 1.

Table 1. Linguistic areas of coverage of WELT Grammar Test (Adapted from Purpura, 2004)

Grammatical Form	Grammatical meaning	Pragmatic meaning (implied)
Phonological or graphological form (e.g. writing system, alphabet); Lexical form (e.g. orthographic form) Morphosyntactic form (e.g. voice, mood, word order); Cohesive form (e.g. adjacency pairs); Information management form (parallelism); Interactional form (e.g. discourse markers)	Phonological or graphological form (e.g. interrogation); Lexical meaning (e.g. style of word) Morphosyntactic meaning (e.g. tense) Cohesive form (e.g. contrast); Information management form (emphasising); Interactional form (e.g. agreeing and disagreeing)	Contextual meanings (e.g. metaphor); Sociolinguistic meanings (e.g. language variation and registers); Sociocultural meanings (e.g. use of apologies, social norms); Psychological meanings (e.g. sarcasm, humour, deference, irony); Rhetorical meanings (e.g. genres).

Two qualifications with regard to the use of Purpura's model are necessary at this point. Firstly, pragmatic meaning cannot readily be assessed through MCT items; however, sentences and longer passages which adhere to accepted social norms, convey particular psychological meanings such as irony or criticism, or represent different rhetorical meanings, may be introduced as positive models within a MCT grammar test. Pragmatic competence is thus promoted through demonstration, using clear models drawn from proficient academic writing. Secondly, although phonological or graphical form or meaning are included in the table, they are not regarded as forming a necessary part of test items in a high-level proficiency test such as this, which already presupposes a sound knowledge of such facets of language.

2. Framework of the study

The remainder of this paper is divided into two main parts. In the first section, BAWE is discussed in terms of aspects of its construction. Three specific themes related to BAWE are investigated: the use of corpora in language testing; the decision to include both first and second language writing (L1 and L2 writing) within the BAWE corpus; and the attributes of proficient academic writing within the BAWE corpus. The second main section provides an overview of the stages involved in developing the WELT MCT test, based on BAWE. This section covers the following areas: the search mechanisms used to locate suitable items; the criteria adopted for including items from the BAWE corpus; and the importance of taking the needs of the test taker into consideration when using

the corpus. This second section focuses primarily on models and examples drawn from BAWE. It leaves out of account considerations of the relative merits and drawbacks of using multiple choice items *per se* (Hughes, 1989; Purpura, 2004), as well as the procedures and pitfalls of devising distracters for the examples harnessed. These are key issues in test design, but they lie beyond the scope of the present study. Finally, a section is provided which discusses aspects of representativeness of particular subject areas and styles of discourse within the newly designed test.

3.   Use of corpora in test development

The use of written corpora in language testing is by no means new. As Barker (2005, 2006) observes, researchers in testing are gradually coming to regard corpora as a key means of enabling testers (as well as stakeholders) to better understand language proficiency ('what happens in tests and why') and to provide appropriate evidence of productive and receptive skills, as well as to enhance and balance other forms of analysis. Testing corpora differ from a straightforward learner corpus, or a corpus such as BAWE, in that they compare a range of student texts produced within the constraints of test/ examination conditions. This type of data remains useful to language testers, since it enables testing specialists to undertake various research procedures that clarify both how the test tasks/questions perform over time, and how test takers are performing.

The present study diverges from Barker's overview by advocating the direct use of student writing to construct items for a language proficiency test. To the best of our knowledge, this is an original feature of the present research. The following three examples, given in Table 2, compare what 'simulated' and 'authentic' examples of particular language areas tested in WELT look like, to show how the sentence stem might differ in each case, albeit testing the same language item.

From the examples given in Table 2, we may see that the naturally occurring examples from BAWE differ significantly from the artificially invented sentences in terms of their linguistic complexity, their enhanced subject-specific focus, and the academic tone and *gravitas* that they convey. In this sense, they clearly convey pragmatic competence through demonstration. There are strengths and limitations in such a corpus based approach. To take the advantages first, one strength is the fact that authentic language (whether at the sentence level or a short passage of two or three sentences) can often be more motivating for test takers, and provoke greater curiosity and interest than is the case with stylised or semi-authentic ones. This is especially the case if the topic areas and genres are seen to be of direct relevance to the test taker, and to come from the discourse community which test candidates will be joining. Whilst Purpura (2004) advances the oft-held view that "context-independent, discrete-point tasks, or those that lack authenticity of topic, are perceived as being…out of touch with their language learning goals." (2004, p. 253), a more positive view is held by Weir (2005, p. 61), who emphasises the possibility of including within a test "texts and

activities that mirror as closely as possible those that candidates ….are likely to meet in their future target situation." Student-generated writing, no doubt, remains closer to the students' experience, and needs, than expert text. Not only this, but items drawn from corpora are more likely to show the use of particular language items (e.g. the present perfect tense, the definite article or a relative pronoun) in naturally occurring contexts (Meyer, 2002; Hunston, 2002). By using a straightforward key word search, or employing appropriate concordancing software, a range of examples for each pre-designated testing point can be called up, the better to undertake an appropriate selection of sentences and longer passages.

Table 2. Comparison between naturally occurring examples from BAWE and artificially invented examples

Sample testing point	Naturally occurring example in BAWE corpus (Test version G1v27)	Artificially invented example (Test version G1v21)
Question form	How does neurobiological development affect cognitive development?	How did he manage to escape?
Present perfect tense	The idea of 'behaviourism' has been round for a long time.	The language laboratory has rarely been used since it was first installed.
Conditional form	In early society, anyone who was good at imitation would have been able to copy the latest hunting skills or stone tool technology.	My friend is going to London next month. I wish I had time to go with him.

Conversely, in using corpus-based information as a source for developing items, there is a risk of selecting a sentence or passage from an assignment where the level of difficulty of the surrounding language is unduly high compared with the item itself. This will impinge on item validity. Furthermore, selecting source material from particular subject-specific assignments (whether it be law, philosophy or mechanical engineering) might inadvertently favour certain test takers over others, especially if they are well versed in these disciplines and have particular world knowledge. For this reason, any item taken from a corpus will need to be adapted and adjusted prior to use. Unconscious bias may readily occur, for instance, when selecting examples, with regard to gender, ethnicity and socio-economic grouping, amongst others. A further difficulty in using items for corpora in a language test might be that of copyright. For this reason, any contributions to a corpus of student writing must be freely useable from a copyright point of view, and contributors need to be made aware of this when they submit their work. Finally, the philosophical question is raised as to whether texts within the BAWE corpus, especially those produced by second language writers, can present suitable "models" for WELT candidates when undertaking multiple choice

grammar items. Whilst we believe that such texts can do so, this key issue will be further discussed in the next section.

4.   First and second language writing: an artificial separation?

A noteworthy feature of the BAWE corpus is that it includes both first language (L1) and second language (L2) writing, though the predominance of the corpus is towards the former. Prior to examining how BAWE can lend itself to test design, it is necessary to examine some of the arguments in favour of, and against, bringing together L1 and L2 writing "under one roof" within the same corpus. In doing so, we should remember that BAWE makes no claim to be a "learner corpus" in any sense, and we may assume that all contributors to the corpus are regarded as proficient university writers, regardless of their first language or geographical background.

According to one school of thought, L1 and L2 writing differ in nature, to the extent that L2 writing models may be regarded as less useful when seeking extracts and models for language tests. This argument is based on the view that L1 writers have greater linguistic competence than L2 writers. As Frankenberg-Garcia (2003) has indicated, language difficulties may exert a negative effect upon the writer's ability to shape their texts, given that the writer will be likely to be preoccupied more by the need for linguistic correctness at the time of writing than anything else. This preoccupation is liable to detract from the writer's ability to engage in higher order tasks. Frankenberg-Garcia shows that L2 writers may also be more subtly influenced by the actual discourse conventions of their L1, so that even if they have good language proficiency, their writing may still "look", and "feel" different from L1 writing, and they may continue to engage in the "negative transfer" of discourse patterns from L1 to L2.

In his comprehensive overview of the issue, Silva (1993) has investigated in detail the distinction between L1 and L2 writing, by examining 72 research studies which have considered this topic, and interviewing 27 writers in detail. Along with other commentators, he sees clear differences between L1 and L2 writing, not only in terms of the process that writers adopt, but the nature of the actual text produced. Silva sees differences, for example, between L1 and L2 writers in the area of "reader orientation". He argues that L2 writers' orientations are longer, and contain a smaller range of attention-getting devices. He also observes, in terms of "morphosyntacticstylistic features", that L2 writing is sometimes less complex, less mature and stylistically appropriate, and less consistent and academic with regard to language, style, and tone. Gilquin and Paquot (2007) further note the tendency of L2 writers to incorporate a greater number of spoken features into their writing than their L1 counterparts.

Other studies avoid separating L1 and L2 writing. Some studies in this vein involve EFL students at an earlier stage of language development, and are of less relevance to this study (e.g. Hirose, 2003). Others tend to be based on contexts that are relatively remote from British higher education, though they demonstrate a wariness of generalising findings. However, some studies emphasise the link between L1 and L2

writing more strongly. For instance, Berman's (1994) research into 126 secondary school students in Iceland, studying English as a foreign language (EFL) shows that while grammar remains central to the transfer of writing skills of the sample group from L1 (Icelandic) to L2 (English), such findings do not help to anticipate the development of academic skills in the future. A further study by Zainuddin and Moore (2003) advises against over-generalising findings from research into the distinction between L1 and L2 writing. The four Malay bilingual writers observed cannot be viewed in terms of any "monolithic cultural entity". Rather, it is found that authors actually tend to vary in terms of their proficiency levels, educational experiences and use of audience-related strategies.

In order to compare the quality of L1 and L2 writing, it is instructive to consider a range of extracts containing the same target item, filtered through the BAWE search engine. A search of the verb "reveals", for instance, produces, amongst others, the following search results:

a) "Website 2 thus *reveals* that these goals may be unrealistic, as achieving them would require certain developing countries to out-perform even the historical rates of progress of the rich countries today". (#0139f – Economics – Methodology Recount - Non-native Speaker).

b) "A literature survey of C [14] dates *reveals* that it happened largely in post Neolithic times in Northern Europe on different former brown earth forest soil sites." (#6019a – Biological Sciences – Explanation – Non-native Speaker).

c) "However the continued over-capacity of the steelmaking industry and price pressures that led to the 2001 crisis *reveals* the strategic incapability of Corus to prosper, and create new opportunities". (#6087f - Agricultural Sciences - Critique - Native Speaker).

d) "Although the second-order broadening is still present in the direct dimension, taking a projection along a certain angle, orthogonal to the ridges in the spectrum, *reveals* a spectrum containing only isotropic shifts". (# 0311i - Physics – Research Report – Native Speaker).

In considering the above extracts, alongside others, few obvious linguistic deficiencies in the L2 samples were observed. Whilst it was important not to over-generalise this, to assume that complete assignments were of totally consistent quality from start to finish, it was nonetheless concluded that both represented the writing of proficient university students

Should both L1 and L2 texts, then, be used as models for test candidates? Our answer here is that excluding L2 writers from a corpus of student writing, and tests, may well lead to a skewed picture of academic writing in British HE institutions, and may

run counter to the aims of the BAWE corpus in terms of its mission to be inclusive. This is not least because international students account for a high proportion of students within UK settings, both at postgraduate, and particularly at undergraduate levels. Neither can one be sufficiently convinced that differences in texts should be attributed purely to the fact that a writer was either L1 or L2, since many other factors shape the writing process: these can include age, gender, educational experience and psychological factors, amongst others. Most of all, it is recognised that the writers who submitted their assignments to the BAWE corpus are deemed to be proficient writers, having obtained good marks for their work. This does not suggest that amendments are not needed before transforming an extract from BAWE into a grammar test item. However, amendments are likely regardless of whether a writer is L1 or L2.

5. Attributes of proficient student writing in BAWE

If student writing is to be used in a language test, the quality of the extracts must be assured. The attributes and qualities of such texts thus require full prior consideration. There is no single approach to measuring the attributes of "good" academic writing, given the inherently subjective nature of judgements. One means of doing so is to consider the type of advice given in manuals of writing skills development. In his well-used guide to essay writing, for example, Roberts (1997) views the elements of proficient writing as constituting the effective writing of sentences; paragraph construction; the production of written summaries; and appropriate referencing skills, amongst others. Many other manuals of this kind exist. Jordan (1997) regards effective academic writing as the achieving of an understanding of the rhetorical-functional approach to writing (including functions such as "instruction", "explanation" and "definition"), coupled with the incorporation of a "process" approach, which incorporates aspects such as self-evaluation, planning and revision/rewriting, amongst others. Crème and Lea (1997, p. 34) also list a number of key elements of academic writing at university level, ranging from "developing an argument" to "linking theory and evidence" and "providing evidence to support an argument." The "use of primary texts" and "using personal interpretation" also feature within this list. Such attempts to define the qualities of proficient academic writing are limited, because they cannot fully take into account the plural, multi-faceted nature of writing undertaken by a range of students. Crème and Lea (1997) concede that the decision as to which of the above elements to incorporate forms one of the chief challenges of academic writing for students: "tutors will have their own understanding of what constitutes a good piece of student writing." (p. 39).

The strengths of the writing contained in the BAWE corpus may be more effectively explained by interviewing lecturers and tutors in the departments where the contributors are based. Undoubtedly, it is the combination of knowing what the features are that subject lecturers require, and being able to identify such features in specific pieces of writing, that assures the quality of a given extract. Nesi and Gardner

(2006) report on previous research in which 55 interviews were conducted with subject lecturers, across 20 departments at the Universities of Warwick, Reading and Oxford Brookes (UK). As the authors comment, there is relative consistency across the interviews with subject lecturers in terms of the importance of "clarity" and "coherent structure". Second to the notion of "clarity", they argue, is the need for "originality" and "creativity". For the purposes of the present study, not all 55 interviews previously undertaken were re-examined; rather, a smaller sample of 25 interviews, from the same interview project, was analysed in detail. A breakdown of this analysis is provided in Table 3.

Table 3. Cross-departmental interviews with subject lecturers examined as part of the BAWE-WELT Project

Faculty	Department	No. of interviews examined
Arts and Humanities	Philosophy	2
	Theatre Studies	2
Social Sciences	Sociology	3
	Economics	3
	Health & Social Sciences	1
	Law	3
	Business School	3
Physical Sciences	Maths	2
	Physics	2
	Molecular Chemistry	1
Life and Medical Sciences	Psychology	1
	Biological Sciences	2

The interview notes from the original interviews were carefully re-read, and the attributes of good writing were isolated and totalled. Where a particular attribute was referred to in different ways (e.g. "originality" was referred to variously as "spark", "critical independence" and "making the subject their own") these qualities were brought together under an over-arching heading, as is shown in Table 4.

From the above analysis, two particular conclusions may be drawn. Firstly, in UK HE settings, the qualities of independence, critical thinking and originality are highly valued, but these might not necessarily suggest themselves as being attributes of appropriate or natural model sentences or extracts for testing purposes; secondly, for language testing purposes, the attributes of clarity, precision, appropriate grammar and spelling are likely to be of greater importance in identifying suitable models for testing purposes. Particularly relevant attributes to look for, as harnessed from the interview

notes, are "short sentences", "writing in chunks" and "avoiding sloppy style". In the next section, we look at how the content of the WELT grammar test was defined, prior to explaining how the items were developed through the use of the corpus.

Table 4. Ranking of attributes of written work observed by subject lecturers across disciplines

Attribute	Alternative descriptions	Frequency of citation
Originality	'Spark', critical independence, making subject their own.	22
Clarity of expression	Short sentences, clear paragraphs, using 'their' voice	13
Logical development	Well composed, progression, flow	11
Precision	Concision, writing in chunks, simplicity.	8
Appropriate citation	Referencing, avoiding 'regurgitation'	5
Grammar and spelling	Avoiding 'sloppy' language, colloquialisms, etc.	5
Showing comprehension of subject matter	Conveying basic message	4
Engagement with material	Interest in what student is talking about,	4
Awareness of own position	Assumptions about reader's position.	4

6.   Developing the WELT grammar paper

This section now considers three particular aspects of test design: the search mechanisms used to harness appropriate extracts from BAWE; the criteria used for including or excluding items; and the importance of taking the needs of the test taker into consideration.

6.1 Search mechanisms

Potential extracts for the shorter, grammar based questions in the paper (items 1-50) were located through a key word search, using a sketch engine programme, to find examples that contained the target grammatical structures (for example, the third conditional, the past perfect tense, or the use of 'in spite of'). This procedure called up a range of examples, of which one or two were ultimately selected for consideration.[4] They were selected on the basis of how understandable they were likely to be for the test takers, and the way in which they contextualised the target language item. An .xls spreadsheet was then used, to match essays with contextual information (for example, the department in which the assignment was written, the year of study, the assignment title). An electronic concordancing search engine was also used, which allowed filters

to be placed on data, and which related the information to the section of text in which it occurred (beginning, middle or end of the assignment). In addition to searching the main BAWE corpus, the earlier pilot corpus, referred to in the first part of this paper, was also used, since this also contained valuable data regarding good quality assignments. Nevertheless, this process was slower, since no electronic search engine was available.

Care was taken to ascertain that the surrounding language of any extracts identified was natural, and not too difficult for our potential test-takers to understand. A number of experienced tutors provided an expert view by commenting on the chosen extracts. For the longer items, from 61 to 100, the primary consideration was to identify an "inductive" type of information structure, in which the information was organised clearly across the progression of the sentences, from general to more specific. Frequently, a transition was sought from a simple opening statement, in sentence 1, to a further exemplification of the thesis statement, rather than a more specific opening. It was felt that test-takers would need to be "led" gradually into the topic area, since they would have had no prior knowledge of the subject area, and might therefore be operating at a disadvantage to the student who had actually written the assignment. In WELT, candidates are required to read a large number of different extracts, some of which were not connected thematically, so it is important to ensure that the complexity of the items does not overload them. In order to locate suitable examples for the second part of the test, the .pdf files of the complete essays were examined, since this helped to identify sections of the text where the writer was presenting, rather than analysing or evaluating information. The most useful sections, for testing purposes, tended to occur earlier in the assignments, but an attempt was made to include items from a range of different locations.

The selected extracts were then matched against the full essay from which they were taken, in order to obtain a better sense of how the extract itself fitted into the organisation of the whole piece of writing. Extracts were read several times, by expert tutors, prior to the piloting stage, to ensure that they had an appropriate level of difficulty, and to ascertain whether they provided useful examples of the language it was supposed to test. Specialised lecturers in English for Academic Purposes were used for this purpose. Feedback from critical readers was paramount in judging the usefulness of test items and also the surrounding language. It was generally regarded as easier to design shorter items than longer ones, and feedback on the longer items was generally more critical, and indicated a wider range of difficulties.

6.2 Criteria for inclusion and exclusion

In re-designing the test, careful judgement was required to achieve a balance between the need for authenticity and the requirement not to overload candidates with undue verbosity and technical jargon. Reasons for rejection included, variously, the difficulty and level of specialisation of the item and the extent to which it highlighted or contextualised the item that we wanted to test (construct validity). Distracters were not

provided at this stage in the process; they were added later, before the trialling stage, and are not analysed as part of this article.

Criteria for inclusion were the accessibility and clarity of language, and the way in which the content and structure of the item clearly illustrated the target structure that we wished to test. This is measured not so much through any conventional view of how difficult an item is, but by examining how many individuals responded correctly to the particular item in pilot versions of the test (facility index). Examples of items that were included are as follows. The BAWE code number for each essay is provided, as well as the subject area, and the testing point of the item characteristics is underlined.

a) "I will begin this essay by looking at an important implication of the development of medicine, *namely* women's exclusion from this field." (#001c – Sociology - Essay). This extract was targeted because of our interest in the connecting word 'namely' as a means of exemplification. It falls into the category of "interactional form", as indicated by Purpura (2004), and represented in Table 1. Given appropriate distracters, the item was seen to work well, and was regarded as useful.

b) "'Imperialism' was at one time a popular concept that *appealed to* Victorian men in England." (#004a – Essay -Sociology). This extract was selected to test the verb "appealed" and its dependent preposition "to". It falls into the categories of "morphosyntactic form and meaning", as specified in Table 1 above. Although regarded by some as a vocabulary, rather than a grammar item, the context was felt to be clear, since the word "popular concept" served to foreshadow the meaning of the actual verb being tested, thereby giving test takers valuable clues as to the meaning of the item.

c) "There has always been disagreement as to whether 'development' is a series of stages, <u>or</u> whether it is *a constantly moving process*." (#0011a – Psychology - Essay). This extract was selected to assess candidates' ability to isolate the contrastive rhetoric of the sentence, and to supply contrasting information, by making the correct choice. This item connects most closely with the idea of grammatical meaning, in Table 1 above, since candidates will need to ascertain the notion of contrast. The clear organisation of information within the sentence ensures that the item is useful in testing what we wish to test (construct validity).

d) "Four types of racing engine were of interest to this project: Formula, Rally, Sports and Motorcycle racing engines. The specifications of these engines have been summarized in Appendix A and B, so that one can *compare their parameters*." (#0018c – Engineering - Essay). This extract was chosen to test candidates on their ability to complete the purpose clause with the correct information. This type of sentence extends the test of grammatical competence to pragmatic meaning, and is seen to be

useful in that it tests the candidate's ability to complete the sentence on the basis of an appropriate understanding of drawing comparisons.

Conversely, some extracts taken from BAWE were rejected as unsuitable. Importantly, these rejections were made on the basis of the suitability of the extracts as test items, and did not constitute a judgement on the writing itself.

a) "In this essay I shall attempt to take a closer look at the conclusions reached in the book." (#0202c - Economics - Essay). This item was designed to test knowledge of a 'statement of intent' structure, often occurring in academic discourse. However, the collocation "take" plus " a look at" was thought to be unknown or unfamiliar to international test-takers and therefore was felt to constitute an unfair item.

b) "In order to *sum up* the essence of political power, we must identify and explain the power which causes these conditions". (#0135a – Politics - Essay). This was designed to test the idiomatic use of common phrasal verbs. However, the repetition of the word "power" in both parts of the sentence was felt to be potentially confusing for potential candidates.

c) "Descartes, a famous French philosopher, believes that there is a different consciousness with a different identity *in every person.*" (#0026b – Philosophy - Essay). This item was intended to be placed early in the test, to assess the difference between quantifiers (each, every, any, etc). It was, however, regarded as a complex framework for testing this structure at such an early stage in the test. The difficulty was evident through the very low facility index of the item (only 7 out of an initial 48 candidates responded correctly). Whilst we did not interview the students to ask why they had found the item difficult, we also assumed that the inclusion of this philosophical topic and culture bound reference (Descartes) might have added a level of complexity to the item.

d) "The importance of workers to the Chinese government, and indeed their own perceived importance, *should not be underestimated.*" (#0135f – Politics - Essay). This extract did not lend itself appropriately to the assessment of what we wished to assess (i.e. it lacked construct validity). We found that the extract was too brief to provide an inductive approach to the subject, and the ending of the sentence was not sufficiently linked to the first part of the sentence to clarify any underpinning information structure, or make it worthwhile including as a test item.

6.3 Test-taker considerations

A crucial aspect of test design to consider was the demands placed on test takers in the test as a whole. It was essential not to overload potential test takers with disparate or potentially confusing examples of language. Thus, some items were grouped

thematically or conceptually. For example, to help the test takers to produce their best work, the first ten items of the test were drawn either from introductions or conclusions to assignments, to enable the test taker to link items with a common function within their mind-set. Despite the varied subject-specific areas of these extracts, this thematic grouping lessened the demands on the reader at an early stage in the test. Some extracts used for these items were as follows:

a) In this essay, I will look at how racism may be defined in contemporary society. (#0001a – Sociology - Essay).

b) I plan to discuss how the positivist school views statistics. (#3029d – Anthropology – Essay- adapted).

c) The aim of this experiment was to investigate isoelectric points of different molecules. (#0009a – Biological Sciences – Methodological Recount).

d) In conclusion, it can be seen that health inequalities still exist. (#0048a – Medicine – Essay).

e) To conclude, the press media is a volatile political battleground. (#0137d – Politics - Essay).

All of the above examples are seen as concise formulations of statements of intent, and are designed to place the underlined target element into sharper relief.

7. Issues of inclusiveness and representation within BAWE and WELT

A key feature of the BAWE corpus is its inclusive approach to academic writing. The merits of incorporating L1 and L2 writing have already been outlined, and some examples have been given to show the type of extracts that were considered acceptable. A further crucial area of inclusiveness in BAWE involves drawing contributions from a range of universities in the UK which have different levels of tradition. In the UK, certain universities, such as the University of Warwick, are considered to be more "traditional" in terms of the range of academic programmes offered, and the assessment mechanisms used. Former polytechnics which became universities after 1992 (sometimes referred to as 'post-1992 universities) continue to have a more diverse student population in terms of socio-economic grouping and ethnicity, and are noted for their vocational-style and applied degree programmes. A number of courses are common to both types of institution. For example, both the more traditional "red brick" universities and their post-1992 counterparts will have business schools, engineering faculties, literature departments and media studies programmes, amongst others. However, dedicated courses in publishing, nursing or automotive

engineering would be more likely to occur in post-1992 institutions, while a post-1992 university may be less likely to offer degree courses in philosophy or French. Furthermore, post-1992 universities often demonstrate a greater sense of innovation and flexibility in terms of assessment types, with self-assessment and oral assessment playing a stronger role. A key feature of the BAWE corpus is its inclusion of many different types of assignment, and a refusal to consider the "essay" as the sole, or indeed, most ideal way of assessing students on their knowledge.

Proficiency test takers will aim to study in a variety of HE institutions in the UK, given that this type of test is generally recognised by both "red-brick" and post-1992 universities alike. Therefore, it seemed crucial for WELT to reflect the diversity of writing inherent in the corpus, by including as wide a diversity of subject areas as possible. In terms of test revision, this procedure involved drawing extracts from subject areas that may be difficult to test within a general proficiency test such as WELT (physical and life sciences, amongst others), and including items from as wide a range of assignment types as possible, such as critiques, data descriptions, reports and personal reflections.

Table 5. Distribution of subject-specific content within revised test

	Introduction and conclusion items (items 1-10).	Single sentence items (items 11-50)	Multiple sentence items (items 61-100)
Arts and Humanities	1	12	12
Social sciences	7	14	15
Physical sciences	0	4	6
Life and medical sciences	0	6	6
General (undifferentiated)	2	4	1

Table 5 demonstrates the broad balance that was achieved, in the new test paper, between items drawn from Social Sciences and Arts and Humanities assignments. In 2008, the spread of items was revised slightly, to match the needs of test takers who were planning to take social sciences oriented courses. As can be seen from Table 5, it proved more problematic to select items from science and engineering assignments, and in consequence, these subjects were not as strongly represented within the test, although a range of items are nonetheless drawn from these areas. Seven items were also included which are designated as 'undifferentiated'. This means that the subject area they are drawn from will not be immediately obvious to the test taker. Two or

three of the "undifferentiated" items were existing items from earlier tests, which were included within the new test as "yardsticks" (anchor items) to ensure that the new version of the test was not more difficult than the old one. Items 51 to 60 were drawn from one specific social sciences paper and are not included within the table.

8. Conclusion

From the above discussion, we may see that the BAWE corpus, and use of it, has progressed considerably from early discussions of the pilot holdings (Nesi, Sharpling & Ganobcsik-Williams, 2004). From these modest and relatively limited beginnings, a major resource has grown, and the BAWE corpus has moved forward to take its position as a rich medium for the study of academic writing. Not only this, but as this article has shown, BAWE has even influenced the development of a resource such as WELT in productive, creative and often surprising ways, and we believe that it is likely to continue to inform the testing of English for Academic Purposes in the future. For example, where a computer-based test of English is envisaged in the coming years, it may be possible to develop even stronger links between the test's infrastructure and a corpus such as BAWE, possibly through interconnecting databases that inform item development. In such circumstances, the meeting of BAWE and language proficiency testing, rather like passing ships travelling in opposite directions, may well develop from being a mere chance encounter to becoming a more protracted, enduring friendship.

Notes

1. The BAWE corpus was developed at the Universities of Oxford Brookes, Reading and Warwick under the directorship of Hilary Nesi. Corpus development was assisted by funding from the ESRC (RES-000-23-0800).The corpus can be accessed through the Oxford Text Archive (http://ota.ahds.ac.uk) as resource number 2539. It includes text files, a spreadsheet with contextual information, and a corpus manual.

2. More detailed information about WELT is available at: http://www2.warwick.ac.uk /fac/soc/al/staff/teaching/sharpling/sharpling_g/testhandbook

3. Sample grammar and English usage test items may be viewed at: http://www2. warwick.ac.uk/fac/soc/al/staff/teaching/sharpling/sharpling_g/weltpracticearchive/practice

4. Key word searches of the BAWE corpus may also be conducted via: http://bawe-search.coventry.ac.uk/BAWEWebApp

References

Alsop, S. & Nesi, H. (2009) Issues in the development of the British Academic Written English (BAWE) corpus. *Corpora, 4*(1), 74-87. doi: 10.3366/E1749503209000227

Barker, F. (2005). What insights can corpora bring to language testing? Retrieved 10 January 2010 from www.ling.lancs.ac.uk/groups/crile/docs/crile%20lectures/barker0105.doc

Barker, F. (2006). *Research Notes*. University of Cambridge ESOL examinations. Special edition on corpora and language testing.

Beare, S. (2002). Writing strategies: Differences in L1 and L2 writing. Retrieved 30 September 2008 from

http://www.llas.ac.uk/resources/paper.aspx?resourceid=1292#toc_14912

Berman, R. (1994). Learner's transfer of writing skills between languages. *TESL Canada Journal, 12*(1), 29-46.

Crème P & Lea, M. (1997). *Writing at University*. Buckingham: Open University Press.

Frankenberg-Garcia, A. (1990). Do the similarities between L1 and L2 writing processes Conceal important differences? Retrieved 10 January 2010 from

http://anafrankenberg.synthasite.com/resources/FG1990DoTheSimilaritiesAndDifferences.pdf

Gilquin, G. & Paquot, M. (2007). Spoken features in learner academic writing: identification, explanation and solution. In *Proceedings of the Fourth Corpus Linguistics Conference, University of Birmingham , 27-30 July 2007* . Retrieved 12 April 2010 from

http://ucrel.lancs.ac.uk/publications/CL2007/paper/204_Paper.pdf

Hirose, K. (2003). Comparing L1 and L2 organizational patterns in the argumentative writing of Japanese EFL students. *Journal of Second Language Writing, 12*(2), 181-209. doi: 10.1016/S1060-3743(03)00015-8

Hughes, A. (1989). *Testing for Language Teachers*. Cambridge: Cambridge University Press.

Hunston, S. (2002). *Corpora in Applied Linguistics*. Cambridge: Cambridge University Press. doi: 10.1017/CBO9781139524773

Hyland, K. (2008). The British Academic Written English (BAWE) Corpus. *Journal of English for Academic Purposes, 7*(4), 294. doi: 10.1016/j.jeap.2008.10.012

Jordan, R. (1997). *English for Academic Purposes. A guide and resource book for teachers*. Cambridge: Cambridge University Press. doi: 10.1017/CBO9780511733062

Meyer, C.F. (2002). *English Corpus Linguistics: An introduction*. Cambridge: Cambridge University Press. doi: 10.1017/CBO9780511606311

Nesi, H, Sharpling, G.P. & Ganobcsik-Williams, L. (2004). Student papers across the curriculum: Designing and developing a corpus of British student writing. *Computers and Composition, 21,* 439-450. doi: 10.1016/j.compcom.2004.08.003

Nesi, H. & Gardner, S. (2006). Variation in disciplinary culture: university tutors' views on Assessed writing tasks. In R. Kiely, R. Rea-Dickins, H. Woodfield & G. Clibbon (eds.). *Language, Culture and Identity in Applied Linguistics*. (pp. 99-118). London: Equinox Publishing,

Nesi, H. (2008). The form, meaning and purpose of university level assessed reflective writing. In M. Edwardes (ed) *Proceedings of the BAAL Annual Conference 2007*. Scitsiugnil Press: London, UK. Retrieved 12 April 2010 from

http://www.coventry.ac.uk/researchnet/external/content/1/c4/30/09/v1192628796/user/form/meaning_purpose.pdf

Purpura, J. (2004). *Assessing Grammar*. Cambridge: Cambridge University Press. doi: 10.1017/CBO9780511733086

Roberts, D. (1997). *The Student's Guide to Writing Essays*. London: Kogan Page Ltd.

Silva, T. (1993). Toward an understanding of the distinct nature of L2 writing: The ESL research and its implications. *TESOL Quarterly, 27* (4), 657-677. doi: 10.2307/3587400

Weir, C. (2005). *Language Testing and Validation: An evidence-based approach*. Basingstoke, Hants: Palgrave MacMillan.

Zainuddin, H. & Moore, R.A. (2003). Audience awareness in L1 and L2 composing of bilingual Writers. *TESL-EJ, 7*(1). Retrieved 30 Sept 2008 from

http://www.cc.kyoto- su.ac.jp/information/tesl-ej/ej25/a2.html