

The BAWE Corpus and Genre Families Classification of Assessed Student Writing

Nesi, H & Gardner, S

Author post-print (accepted) deposited by Coventry University's Repository

Original citation & hyperlink:

Nesi, H & Gardner, S 2018, 'The BAWE Corpus and Genre Families Classification of Assessed Student Writing', *Assessing Writing*, vol. 38, pp. 51-55.

<https://dx.doi.org/10.1016/j.asw.2018.06.005>

DOI [10.1016/j.asw.2018.06.005](https://dx.doi.org/10.1016/j.asw.2018.06.005)

ISSN 1075-2935

Publisher: Elsevier

NOTICE: this is the author's version of a work that was accepted for publication in *Assessing Writing*. Changes resulting from the publishing process, such as peer review, editing, corrections, structural formatting, and other quality control mechanisms may not be reflected in this document. Changes may have been made to this work since it was submitted for publication. A definitive version was subsequently published in *Assessing Writing*, 38, 2018 DOI:

[10.1016/j.asw.2018.06.005](https://dx.doi.org/10.1016/j.asw.2018.06.005)

© 2018, Elsevier. Licensed under the Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International

<http://creativecommons.org/licenses/by-nc-nd/4.0/>

Copyright © and Moral Rights are retained by the author(s) and/ or other copyright owners. A copy can be downloaded for personal non-commercial research or study, without prior permission or charge. This item cannot be reproduced or quoted extensively from without first obtaining permission in writing from the copyright holder(s). The content must not be changed in any way or sold commercially in any format or medium without the formal permission of the copyright holders.

This document is the author's post-print version, incorporating any revisions agreed during the peer-review process. Some differences between the published version and this version may remain and you are advised to consult the published version if you wish to cite from it.

The BAWE Corpus and Genre Families Classification of Assessed Student Writing

Authors: Hilary Nesi and Sheena Gardner

Emails: h.nesi@coventry.ac.uk ; sheena.gardner@coventry.ac.uk

Affiliations: Coventry University Abstract

Abstract

The British Academic Written English (BAWE) corpus (www.coventry.ac.uk/BAWE) comprises almost 3,000 pieces of university student writing distributed across four domains (Arts & Humanities, Life Sciences, Social Sciences, Physical Sciences) and four levels of study (from first year undergraduate to taught Master's level). The texts had all been submitted as part of regular university coursework, and had been awarded top grades, indicating that they had met disciplinary requirements in terms of level and task. The corpus was compiled to enable identification of the linguistic and generic features associated with successful university student writing. Our detailed analyses of the corpus led to the identification of thirteen genre families, and supports the premises that university students write in a wider variety of genres than is commonly recognised, and that student writing differs across genres, disciplines and levels of university study. This review introduces the BAWE corpus and the associated genre family classification, then explains how they can be accessed and used for teaching and research purposes, how they have been used to deepen our understanding of academic writing in English, and where they have been used to inform the development of online, interactive and paper-based English language teaching materials.

Keywords corpus; student writing; academic English; genre family; genre classification

Highlights

- BAWE is a 6.5 million word corpus of successful student writing in English
- BAWE analysis led to a genre family classification of university student writing
- Metadata describes the text, the writer and the disciplinary context
- BAWE is available for use by researchers, teachers and students.
- Publications describe the corpus and its applications internationally

Tool: The British Academic Written English (BAWE) Corpus (www.coventry.ac.uk/BAWE)

Tool Purpose: To enable identification of the linguistic features associated with successful university student writing

Key Premises: University students write in a wider variety of genres than is commonly recognised. Student writing differs across genres, disciplines and levels (years) of university study.

Research Connections: The British Academic Spoken English (BASE) corpus; the Michigan Corpus of Spoken Academic English (MICASE); the Michigan Corpus of Upper-Level Student Papers (MICUSP).

Limitations: Data from English universities written between 2000 and 2007. Small sample sizes in some disciplines and genre families.

Future Developments: Other academic written English corpora reflecting university writing requirements in other regions of the world (e.g. the Academic Writing at Auckland (AWA) corpus, <https://awa.auckland.ac.nz>).

I. Introduction

The British Academic Written English (BAWE) corpus is a 6.5 million word collection of successful university assignments, written in English. The assignments were mostly collected in 2005 and 2006, from students at the Universities of Reading, Warwick and Oxford Brookes, although there were some additions to the corpus in 2007, including from Coventry University. The collection process, explained in detail in Alsop and Nesi (2009), was informed by our experience constructing a smaller pilot corpus in 2001 and 2002, as described in Nesi et al. (2004). As the pilot project accepted any successful assignment that any student wished to contribute, the disciplines and levels of study were very unevenly represented. For example, assignments written by second and third level undergraduates, and by Humanities students, were easier to obtain than those by first and fourth level students, and those in the hard sciences. Our collection procedures for the BAWE corpus were designed to create a better balance across levels of study and across disciplinary groups (see Table 1), although sample sizes remained small in a few disciplines, most notably Anthropology (49), Publishing (30) and disciplines placed in the 'Other' category, such as Education (9).

The four universities we chose as collection sites represented a range of British university types: Reading University is an older 'red brick' institution, founded in the 19th century, Warwick is a more modern 'plate glass' institution dating from the 1960s, and Oxford Brookes and Coventry are counted as 'post-1992' because they only acquired university status in that year (having previously functioned as polytechnics, with an applied, technical focus). Because of their different histories they offer a complementary range of degree programmes, both pure and applied, across faculties. This enabled us to collect assignments from a wider range of disciplines than would have been possible if we had focussed on a single institution.

The contents of the BAWE corpus are outlined in Table 1. Holdings come from more than 30 disciplines, grouped into four domains: Arts and Humanities, Life Sciences, Social Sciences and Physical Sciences. The levels in this table refer to the first, second and final year of undergraduate study (Levels 1 to 3), and taught masters programmes (Level 4). Some pieces of assessed work contain multiple texts, a portfolio of lab reports, for example, so the corpus actually contains somewhat more texts than assignments.

Table 1. Overview of BAWE Corpus Holdings

		Level 1	Level 2	Level 3	Level 4	Total
Arts and Humanities (AH) *Linguistics, English, Philosophy, History, Classics, Archaeology, Comparative American Studies, Other	Students	101	83	61	23	268
	Assignments	239	228	160	78	705
	Texts	255	229	160	80	724
	Words	468,353	583,617	427,942	234,206	1,714,118
Life Sciences (LS) *Biology, Agriculture,	Students	74	71	42	46	233
	Assignments	180	193	113	197	683
	Texts	188	206	120	205	719

Food Sciences, Psychology, Health, Medicine	Words	299,370	408,070	263,668	441,283	1,412,391
Physical Sciences (PS) *Engineering, Chemistry, Computer Science, Physics, Mathematics, Meteorology, Cybernetics & Electronics, Planning, Architecture	Students	73	60	56	36	225
	Assignments	181	149	156	110	596
	Texts	181	154	156	133	624
	Words	300,989	314,331	426,431	339,605	1,381,356
Social Sciences (SS) *Business, Law, Sociology, Politics, Economics, Hospitality Leisure & Tourism Management, Anthropology, Publishing	Students	85	88	76	64	313
	Assignments	207	197	166	207	777
	Texts	216	198	170	207	791
	Words	371,473	475,668	440,674	688,921	1,999,130
Total students		333	302	234	167	1039
Total assignments		807	767	591	587	2761
Total texts		840	787	602	620	2858
Total words		1,440,185	1,781,686	1,558,715	1,704,015	6,506,995

*Disciplines are listed in decreasing order of number of assignments collected.

All the BAWE assignments are deemed to be successful in that they were written for accredited degree programmes and had been awarded good grades by the students' subject tutors. Grading practices varied across departments and modules, so in the final version of the corpus the original assignment grades were replaced by the letters D ('distinction'), equivalent to 70% or more, and M ('merit'), for assignments graded between 60% and 70% and those in the Medical School which had been graded on a pass/fail basis.

II. Related Research: From the BAWE corpus to Genre Families

BAWE was developed as part of 'An investigation of genres of assessed writing in British Higher Education' (ESRC RES-000-23-0800 2004-2007). This project aimed to identify and describe genres of student writing, combining large-scale corpus-based analyses with more detailed examination of linguistic features (Nesi et al.2008). Its creation was motivated by the recognition that, at the time, very little was known about many of the genres university students produce. University assignment genres are in fact highly "occluded" (c.f. Swales, 1996), and only rarely enter the public domain (dissertations and theses, in contrast, are often accessible). Lecturers and subject tutors see the assignments they assess, but not necessarily those assessed by their colleagues; students usually only get to see their own efforts (and perhaps those of a few friends); writing teachers and writing assessors are more likely to encounter unsatisfactory, low-grade assignments than those that meet departmental expectations and conform to disciplinary conventions. Admittedly some large-scale

surveys of assignment types had been conducted prior to the creation of the BAWE corpus (c.f. Gardner & Nesi 2013), but although these led to task classifications, they relied on information provided by academics (e.g. Rosenfeld et al. 2004), or course documentation (e.g. Moore & Morton 2005) and were not informed by examination of the students' own writing. The interviews with staff undertaken as part of the BAWE project provided helpful information regarding disciplinary expectations, but in our experience the nomenclature used by our informants was often unreliable, and many different types of writing were described as either 'reports' or 'essays', often interchangeably and without acknowledging that different assignments with the same descriptor might require different uses of language and different organisational patterns. Student contributors to BAWE tended to share this lack of awareness, and although the collection process required them to choose (from a drop-down menu) the type of assignment they had submitted, it was common for work submitted as a 'report' to be described as an 'essay' within the assignment itself, and vice versa. The genre classification therefore was not reliant on the participant nomenclatures; rather texts were analysed as genres; using Martin's definition of genre as a 'system of staged goal-oriented social processes through which social subjects in a given culture live their lives' (1997: 13).

Because the BAWE corpus contains a large amount of proficient writing produced in response to a wide range of assignment briefs, we were able to identify similarities and differences in the language and organisation of assignments, whether or not they were described well by staff, students or the assignment brief. We read all the student texts in the corpus several times over, and gradually constructed a list of 13 'genre families', each containing diverse genres, as described in Table 2. These families included assignments from different disciplines and levels of study, but with similar social purposes, and with certain family resemblances in terms of language functions and organisational structure.

Table 2. BAWE Genre Families and their Purposes

Social purposes	Genre family	Examples of genres
Demonstrating knowledge and understanding	Exercise	<i>calculations; data analysis;. calculations + short answers; short answers; statistics exercise</i>
	Explanation	<i>legislation overview; instrument description; methodology explanation; site/ environment report; species / breed description; account of a natural phenomenon</i>
Developing powers of independent reasoning	Critique	<i>academic paper review; interpretation of results; legislation evaluation; policy evaluation; programme evaluation; project evaluation; review of a book/ film/ play/ website</i>
	Essay	<i>challenge; commentary; consequential; discussion; exposition; factorial</i>
Building research skills	Literature Survey	<i>annotated bibliography; anthology; literature review ; review article</i>
	Methodology Recount	<i>data analysis report; experimental report; field report; forensic report; lab report; materials selection report</i>
	Research Report	<i>research article; research project; topic-based dissertation</i>
Preparing for professional practice	Case Study	<i>business start-up; company report; organisation analysis; patient report</i>
	Design Specification	<i>building design; game design; product design; website design</i>

	Problem Question	<i>law problem question; logistics simulation; business scenario</i>
	Proposal	<i>book proposal; building proposal; business plan; catering plan; marketing plan; policy proposal; research proposal</i>
Writing for oneself and others	Empathy Writing	<i>expert advice to industry; expert advice to lay person; information leaflet; job application; letter; newspaper article</i>
	Narrative Recount	<i>accident report; account of literature or website search; biography; creative writing: short story; plot synopsis; reflective recount</i>

The distribution and nature of assignments across the genre families and disciplines is described in detail in Nesi and Gardner (2012). With one exception (Problem Questions in Arts and Humanities) every genre family was found in all four disciplinary domains. It is therefore likely that knowledge of their features and organisational patterns would benefit most students, regardless of their disciplinary background.

III. Connections

Gardner (2016) provides comprehensive advice on the use of the BAWE corpus to develop discipline-specific teaching and assessment materials, and explains the relationship between five linked resources:

- (1) published research that investigates assignment genres and registers;
- (2) BAWE spreadsheet descriptions of the corpus contents;
- (3) the BAWE corpus itself;
- (4) online teaching materials based on the above; and
- (5) lesson plans from EAP teachers in pre-sessional and in-sessional contexts.

The value of these five resources is demonstrated with specific reference to Business Case Studies, Economics Essays and Engineering Methodology Recounts, assessment contexts for a high proportion of international students. Gardner illustrates how information from each resource can be extracted and interpreted.

III.ii Research on the nature of successful assessed student writing

The Coventry BAWE website (www.coventry.ac.uk/BAWE) lists dozens of articles drawing on BAWE data and employing a mix of theories and methodologies to develop our understanding of academic English. Some of these articles specifically address questions teachers or students might ask, such as 'Can I use section headings in my essay?' (Gardner & Holmes 2009), and many include examples of classroom teaching and testing activities designed to assess specific features. For example, Lee and Chen (2009) explore high frequency verbs in the corpus (make, do) while Nesi and Moreton (2012) examine students' use of shell nouns. The successful student writing referred to in these articles provides a more appropriate model for student writers than, for example, professional writing in textbooks or journals.

III.iii Accessing the BAWE corpus for research and teaching

As explained on the BAWE website, the corpus is available for research purposes (e.g. to inform assessment) through the Oxford Text Archive (OTA). As explained on the OTA website (<https://ota.ox.ac.uk/>), this Archive "develops, collects, catalogues and preserves electronic literary

and linguistic resources for use in Higher Education, in research, teaching and learning". The archive also advises on resource creation and use, and "the development of standards and infrastructure for electronic language resources".

Classics	illustrates the decline of the hetaira, Lais, when he contrasts the period of her prime, when starters (two drachmas) made
Linguistics	of tense. </p><p> This essay firstly defines the term 'tense' and contrasts the theoretical approaches of structuralists and logical
Engineering	considers the benefits and limitations of these tools and contrasts them manual techniques such as wind tunnel testing and
Philosophy	example he gives poetic descriptions of Death and Hell and contrasts these to paintings of the same subjects. The paintings, in
English	such as Pope, Swift and Sheridan. In the Preface Wordsworth contrasts these two types of language describing the latter as "
Sociology	to the public is of a scandalous nature. Bainbridge contrasts this view of the media with that of researchers who are more

Figure 1. SketchEngine display of concordance lines for contrasts in BAWE

The BAWE corpus can be freely accessed by students, teachers and test designers through the open SketchEngine, as illustrated in Figure 1. A guide to using SketchEngine with BAWE (Nesi and Thompson n.d.) is available to download, and the BAWE QuickLinks project <http://bawequicklinks.coventry.domains> offers tutors ways of linking their feedback to SketchEngine concordance lines. Other interfaces that enable access to BAWE data include the well-known LexTutor site.

III.iv Materials in the public domain informed by BAWE and the Genre Family Classification

The most widely available materials based on BAWE data are those produced in collaboration with the British Council for the open-access Writing for a Purpose project (ESRC ES/J010995/1 2012-3) (<https://learnenglish.britishcouncil.org/en/writing-purpose/writing-purpose>). Taking a specific genre family, discipline or writing purpose as their starting point, users of these materials can access texts and activities which promote language development and an understanding of academic writing. The language 'resources' section not only links to word lists for disciplines and levels of study (<http://www.uefap.com/writingforapurpose/vocabulary/wordlists/index.htm>), available on cowriter Andy Gillet's UEFAP website, but also leads to 'useful vocabulary' for each genre family, illustrated in context via links to the SketchEngine site.

Although students might need some assistance to navigate Writing for a Purpose, it contains a wealth of information for teachers wishing to design tests and materials focussing on genres, disciplines and lexico-grammatical features relevant to authentic university assessment contexts.

The BAWE corpus genre family classification has also provided a theoretical underpinning for online materials in Hong Kong. The Literacy in the Disciplines (LID) project developed at Hong Kong Polytechnic University, City University of Hong Kong and the Hong Kong Baptist University (<http://literacy.elc.polyu.edu.hk/>) is noteworthy here for its specific links to assessment, as can be seen in the genre assessment checklists (<http://literacy.elc.polyu.edu.hk/socialscience/student/genre-checklist>) which are informed by detailed genre analysis, developed jointly by content lecturers and English language tutors, and used for assessment with students. This excellent practice extends the current genre descriptions available for BAWE to the Hong Kong context. Research for the LID project is ongoing, with additional genres being added to the website as materials are developed.

While the LID project is concerned with undergraduate writing across disciplines, the Online Support for Academic Writing for the Engineering Curriculum (OSAWEC) project (<https://osawec.elc.cityu.edu.hk/repo/>), at City University of Hong Kong, the Hong Kong Polytechnic University and the University of Hong Kong, focuses more specifically on postgraduate genres in Engineering, ranging from journal articles to PhD reports and theses

In addition to these online materials, paper-based resources have also been informed by the BAWE corpus and the genre family classification. These include the Oxford Learner's Dictionary of Academic English (2014) which has detailed sections on writing essays, case studies, reflective writing and reports; Garnet's 50 Steps to Improving Your Academic Writing (Sowton 2012); and Essentials of Essay Writing: What markers look for in the Palgrave Study Skills series (Roberts 2017). Finally, the corpus and genre family classification are widely referenced internationally in teacher education programmes and resources (e.g. Charles and Pecorari, 2015)

IV Possibilities and Limitations

Future work with the BAWE corpus might take a number of directions. Developments in corpus tools and technologies are likely to enable the corpus to be used with increasingly varied interfaces, offering more opportunities for comparisons with other corpora, more ways to conduct analyses, and simpler corpus access routes for busy teachers and students.

Of course, there is still much that we do not understand about the registers of assessed student writing. Although it has long been known that attention should be paid to features such as evaluation, heteroglossia, and complexity, studies such as those of Staples et al. (2016), Nesi and Gardner (2017) and Gardner, Nesi and Biber (2018) are only beginning to unpick how these are differentiated across genres, disciplines and levels of study. Additionally, there is scope to enrich the BAWE genre family classification with reference to a wider range of contexts, both internationally and in terms of disciplines, genres and levels of study. The theory implicit in the classification is that the five broad social purposes and the thirteen genre families can account for assessed university student writing in all contexts, but this is open to empirical investigation. Since the data was collected, a greater variety of digital tools and technologies have come into use for assessment and feedback, and the outcomes of these new approaches may challenge or lead to modifications to our classification scheme.

We hope that there will be more teaching and assessment materials based on the BAWE corpus and its genre family classification. We know there are still EAP programmes that predominantly teach essay writing to students who may never be asked to write an essay in their academic programmes, and we know that some writing tutors still fail to distinguish assignment types beyond the 'essay' and the 'report'. We eagerly await more LID-type assessments that bring together language experts and subject experts, and present to students a shared and mutually supportive view of academic writing.

References

- Alsop, S. & H. Nesi. (2009) Issues in the development of the British Academic Written English (BAWE) corpus. *Corpora* 4(1) 71-83
- Charles, M. & D. Pecorari. (2015) *Introducing English for Academic Purposes*. Abingdon: Routledge.
- Gardner, S. (2012) 'Genres and registers of student report writing: an SFL perspective on texts and practices' *Journal of English for Academic Purposes* 11(1) 52–63.
<https://doi.org/10.1016/j.jeap.2011.11.002>
- Gardner, S. (2016) A genre-instantiation approach to teaching English for Specific Academic Purposes: Student writing in Business, Economics and Engineering. *Writing and Pedagogy* 8(1) 117-144. DOI: 10.1558/wap.v8i1.27934.
- Gardner, S. & H. Nesi. (2013). A classification of genre families in university student writing. *Applied Linguistics* 34(1) 1–29, <https://doi.org/10.1093/applin/ams024>

- Gardner, S. & J. Holmes. (2009) *Can I use headings in my essay?* Section headings, macrostructures and genre families in the BAWE corpus of student writing In Maggie Charles, Susan Hunston, Diane Pecorari (Eds) *Academic Writing: At the Interface of Corpus and Discourse*. London: Continuum. pp. 251–271. ISBN 978-1-84706-436-3
- Gardner, S., Nesi, H. & D. Biber. (2018) Discipline, level, genre: Integrating situational perspectives in a new MD analysis of university student writing. *Applied Linguistics*.
<https://academic.oup.com/applij/advance-article/doi/10.1093/applin/amy005/4937797?guestAccessKey=137ec598-1e45-4428-ae5f-047b7d47e4e2>
- Lea, D. (ed.) (2014) *Oxford Learner's Dictionary of Academic English*. Oxford: Oxford University Press
- Lee, D., & X. Chen. (2009). Making a bigger deal of smaller words: function words and other key items in research writing by Chinese learners. *Journal of Second Language Writing*, 18(4), 281–296.
- Lughmani, S. D., Gardner, S., Chen, J., Wong, H., & L. Chan. (2016) English across the Curriculum: Fostering collaboration. *ELTWO* Retrieved from https://blog.nus.edu.sg/eltwo/files/2016/12/2-English-Across_Lughmani-et-al-221216-1n0bntj.pdf
- Martin, J.R. (1997) 'Analysing genre: functional parameters' in Christie, F. and Martin, J.R. (eds.): *Genres and Institutions: Social Processes in the Workplace and School*. London: Continuum.
- Moore, T. & J. Morton. (2005) Dimensions of difference; a comparison of university writing and IELTS writing. *Journal of English for Academic Purposes* 4(1) 43-66
- Nesi, H., & E. Moreton. (2012). EFL/ESL writers and the use of shell nouns. In R. Tang (Ed.), *Academic Writing in a Second or Foreign Language: Issues and challenges facing ESL/EFL academic writers in higher education contexts*, pp. 126–145. London: Continuum.
- Nesi, H. & S. Gardner. (2017) 'Stance in the BAWE Corpus: New Revelations from Multidimensional Analysis' Corpus Linguistics 2017 Conference, University of Birmingham, 25-28 July, 2017.
<http://www.birmingham.ac.uk/Documents/college-artslaw/corpus/conference-archives/2017/general/paper257.pdf>
- Nesi, H. & S. Gardner (2012) *Genres across the disciplines: Student writing in Higher Education*. Cambridge: Cambridge University Press.
- Nesi, H., Gardner, S., Thompson, P., Wickens, P., Forsyth, R., Heuboeck, A., Holmes, J., Hindle, D., Ebeling, S., Leedham, M., & S. Alsop. (2008) *An Investigation of Genres of Assessed Writing in British Higher Education: Non-Technical Summary (Research summary) ESRC End of Award Report*, RES-000-23-0800. Swindon: ESRC. <http://www.researchcatalogue.esrc.ac.uk/grants/RES-000-23-0800/outputs/read/4cd709c1-dd78-4936-8edf-6a91e384c8d5>
- Nesi, H, Sharpling, G. & L. Ganobcsik-Williams. (2004) The design, development and purpose of a corpus of British student writing. *Computers and Composition* 21(4) 439-450
- Nesi, H. & P. Thompson. (nd) Using SketchEngine with BAWE.
<http://www.coventry.ac.uk/research/research-directories/current-projects/2015/british-academic-written-english-corpus-bawe/search-the-bawe-corpus/>
- Roberts, J.Q. (2017) *Essentials of Essay Writing: What markers look for*. London: Macmillan Education/ Palgrave.
- Rosenfeld, M., R. Courtney & M. Fowles. (2004). *Identifying the Writing Tasks Important for Academic Success at the Undergraduate and Graduate Levels*. GRE Board Research Report No. 4. Educational Testing Service.
- Sowton, C. (2012) *50 Steps to Improving Your Academic Writing*. Reading: Garnet Education.

Staples, S. Egbert, J. Biber, D. & B. Gray. (2016). Academic writing development at the university level: Phrasal and clausal complexity across level of study, discipline, and genre. *Written Communication*, 33(2) 149–183.

Swales, J. (1996). Occluded genres in the academy: the case of the submission letter. In E. Ventola & A. Mauranen (Eds.) *Academic writing: Intercultural and Textual Issues*, pp. 45–58. Amsterdam: John Benjamins.