Sources for courses: metadiscourse and the role of citation in student writing

Nesi, H

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

Original citation & hyperlink:
'Sources for courses: metadiscourse and the role of citation in student writing', Lingua, vol. 253, 103040, pp. (In-press).
https://doi.org/10.1016/j.lingua.2021.103040

DOI 10.1016/j.lingua.2021.103040
ISSN 0024-3841

Publisher: Elsevier

© 2021, 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.

Sources for courses: metadiscourse and the role of citation in student writing

Hilary Nesi (2021) *Lingua* Special Issue on Metadiscourse.

Abstract

Although a great deal has been written about citation in expert research writing, and about novice students' acquisition of citation skills for the purposes of argumentation, little is known about the typical uses of citation in undergraduate student coursework, in different disciplines, and at different levels of study. This paper describes how typical citations patterns can be retrieved from undergraduate assignments in the British Academic Written English (BAWE) corpus, and examines the linguistic features of these patterns with reference to the types of sources cited and the different functions of citation. It reveals a wide variety of patterns and purposes, well beyond those described in student writing guides. Students' use of citation was found to increase as they progressed through the years of undergraduate study, although citation was not an essential feature of every type of writing assignment, and was performed in different ways according to discipline. The search queries presented in the paper can be used by researchers, students, and writing tutors as a basis for further investigation into citation practices. The findings also have implications for teaching, and the training of EAP practitioners.

Keywords

BAWE; citation; student writing; source use; genres

1. Introduction

Citation (writers' acknowledgement of outside sources) is of particular interest to applied linguists because the skills needed to cite meaningfully, accurately and appropriately are difficult to acquire, and because citation is one of the main means by which academic writers signal knowledge of their field of study, membership of their discourse community, and alignment with particular schools of thought or practice. Although not unknown in professional texts outside academia, most professional writers do not need to acknowledge their sources in the explicit manner of academics. This may be because much of their writing is addressed to non-experts, or experts in a different field, and generally these readers trust to

the writer's professional expertise and accept that evidence exists for the writer's claims. Thus, because it is a peculiarly academic practice, most students enter university with very little understanding of the purposes or the mechanics of citation; conventions and skills are expected to be acquired through exposure to academic texts, in academic writing classes, and through the guides which are readily available in university libraries and writing centres.

In Systemic Functional Linguistics (SFL), as described by Halliday (1994), the ideational, interpersonal and textual metafunctions play a central role. Although according to SFL theory each metafunction is manifested in every act of language use, the three metafunctions can be treated as a way of organising the different aspects of citation practice. Xu (2016), for example, associates writers' choice of source to the ideational metafunction (as sources must be appropriate to the field and the topic of the text), and associates the mechanics of citation and the writer's choice of referencing style with the textual metafunction. Libraries and university writing centres often seem to focus on the ideational and textual aspects of citation, providing lists of recommended texts and referencing style manuals. Student writing guides tend to present citation as a means of avoiding the accusation of plagiarism and emphasise the importance of providing accurate and consistent information about sources. The *Purdue Online Writing Lab* (undated) provides 'Preventing Plagiarism' pages, for example, and Marggraf Turley (2016: 80), in his book *Writing Essays: A Guide for Students in English and the Humanities*, stresses the need to avoid ambiguity: "the whole point of giving clear accurate references is to enable your readers to follow up for themselves any interesting books or articles mentioned in your discussion". This is in fact an unlikely primary purpose for citation, however. As Swales (2014:121) points out, "it is to be doubted if article authors (excluding authors of review articles) often have this function firmly in mind".

Student writing guides are less likely to advise on other textual aspects of citation, such as tense choice, as discussed by Hawes and Thomas (1997), or the differing functions of integral and non-integral citation, as discussed by Swales (1990). On the other hand most applied linguistic studies of citation are concerned with its interpersonal role, often regarded as inseparably interwoven with its textual functions (Hyland and Tse 2004).

The interpersonal and textual functions of citation are a rich area for linguistic investigation, and have been acknowledged within several influential metadiscourse studies. In Hyland's framework (2005), for example, citation falls in to the "Evidential" category of interactive metadiscourse features; resources in this category indicate "the writer's awareness of a

participating audience and the way he or she seeks to accommodate its possible knowledge, interests, rhetorical expectations and processing abilities" (Hyland. 2005, p. 49). This makes the point that writers choose their sources and the form of their citations to suit the requirements and interests of their readers. In this regard, Pecorari (2006) is concerned with the way in which writers reveal to their readers the relationships between their sources and the citing texts. Readers have no option but to rely on the interpretations offered by writers, as they are unlikely to know all the source texts sufficiently well to work out these relationships for themselves. Citation thus entails a 'tacit promise' to the reader that it has been carried out in accordance with the conventions of the relevant discourse community (Pecorari 2006: 6). Of course in the case of student writers the primary reader is the lecturer who will grade the text, and who will want to see "authorial demonstrations of due diligence" (Swales 2014: 119) as well as the ability to interpret sources and connect them appropriately. Citation can also be considered metadiscoursal in Vande Kopple's terms, although it is not explicitly included in his taxonomy of metadiscourse (2002, 2012). Vande Kopple describes metadiscourse as a means of helping readers to "connect, organize, interpret, evaluate, and develop attitudes" towards text (Vande Kopple 2012:38). Citation always realizes at least some of these functions: at a basic level it connects the current text with earlier sources, but it can also make connections between earlier sources, position the writer in relation to these sources and indicate the writer's attitude and level of commitment to the propositions these sources express.

The meanings and uses of reporting verbs have been categorised extensively by Thompson and Ye (1991), Hunston (1993, 1995), Hyland (1999) and Charles (2006). Hyland (1999) distinguished between verbal processes representing acts of 'discourse', 'cognition' and 'research', and Charles (2006) offered a roughly equivalent categorisation of reporting verbs according to four central purposes: ARGUE, THINK, SHOW, and FIND. Hyland (1999) also considered the role of the reporting verb as an indication of a writer's acceptance or rejection of a source proposition, and of the original author's acceptance or rejection of an earlier claim. In this respect Hyland's work overlaps with the Engagement domain within Appraisal Theory (Martin and White 2005) which is concerned with a broader range of evaluative features. Martin and White build on Bakhtin's (1981) concept of 'heteroglossic' text which signals the possibility of alternative viewpoints, adding the contrastive phenomenon of 'monoglossia' to describe text which asserts the 'facts' without making them open to debate. In Martin and White's Engagement domain the Endorse category signals that the writer accepts

that a reported proposition is correct (using a reporting verb such as *show*, *prove*, *demonstrate*, *find*, or *point out*), while Acknowledge conveys a neutral position, for example through the use of the adjunct *according to* or reporting verbs such as *report* or *suggest*, and Distance signals that the writer does not take responsibility for what is being reported, for example through the use of a reporting verb such as *claim.* Reporting signals (verbs or adjuncts) always signal some degree of heteroglossia, but non-integral citations enable writers to make bare, unhedged, monoglossic assertions which are not presented as being negotiable. Citation is always metadiscoursal, however, regardless of the form it takes, because it always has a connective function, entails some sort evaluation of the source (albeit implicit), and intends to accommodate the reader's communicative needs.

There is a large body of applied linguistic research examining source use and citation in English academic texts. This has looked on practice from various perspectives, such as disciplinary culture (e.g. Charles 2006, Harwood 2009) national culture (e.g. Hu & Wang 2014), the writer's first language (e.g. Cumming et al 2018) and/or the writer's level of expertise (e.g. Mansourizadeh & Ahmed 2011, Swales 2014). Swales' study is of particular relevance because it drew on the Michigan Corpus of Upper Level Student Papers (MICUSP), a similar corpus to the British Academic Written English (BAWE) corpus used in the present study. Swales compared the work of final year undergraduate biology students with their graduate counterparts, but found little difference in their citing behaviour. He was able to describe and quantify structural features and reporting verb choices, and as he was only working with 37 texts he could analyse their attempts at evaluative engagement in some detail. Nevertheless he was very wary about ascribing citation function on the basis of purely textual evidence, considering it to be "a subjective and rather chancy business" (2014: 120). He goes on to point out that regardless of the research method, as even in interview studies such as those of Harwood (2008, 2009), informants were not always able to explain the purposes of their citations. In the present study I infer the citation forms best adapted to various citation purposes, drawing on the prior literature and on textual evidence. Relevant insights are reported within the results section, but purposes are not quantified or correlated to form in any systematic way, and given the still inexpert status of the BAWE contributors we must assume that these purposes were not always expressed well, and the relevant discourse conventions were not always followed.

Coming from a background in education rather than applied linguistics, Bizup (2008) developed a way of classifying sources according to the use that students make of them, aiming "to name, describe, and analyze the different ways writers use their materials on the page or, equivalently, the various postures toward their materials that writers adopt" (2008: 75). He identified four types of sources, according to whether they are used for the purposes of 'Background', 'Exhibition', 'Argument' or 'Method'. Bizup's Argument sources are those which writers interact with and discuss, while Background sources are ones that writers treat as authoritative and not open to debate. Exhibit sources are there for writers to analyse or interpret, perhaps in multiple ways if the source is rich, and Method sources supply writers with procedures, theoretical models, or terminology. Bizup proposed this classification as a framework for reading, to help students recognise how complex texts deploy multiple sources of different types. He did not discuss the effect of genre or discipline, but it is easy to see how such considerations might influence source type choices, and how these choices might in turn influence the linguistic means by which citation is expressed.

Indeed, the discourse functions of citations identified by Petrić and Harwood (2013) ("position", "defining", "supporting", "application", "agreement", "disagreement", "topic relevance" and "acknowledgement") share some of the purposes identified by Bizup. Background sources would be needed to define, and demonstrate topic relevance, Method sources would be needed for the application of procedures and theories, while sources in Bizup's Argument category would be needed for agreement and disagreement, and for supporting citations, which "help justify the writer's claim or idea" (Petrić and Harwood 2013:114). Cumming et al (2018) later treated Petrić and Harwood's "position", and "acknowledgement" functions as part of a "supporting" category, as they too seem to suit citations which are used for the purposes of argument.

Applied linguistic research into academic discourse emphasises the role of argument, and studies of citation in relation to writer stance have tended to draw their data from argumentative texts, such as research articles (e.g. Hyland 2002; Hood 2010, Mansourizadeh & Ahmad 2011, Silver 2012, Author and other 2019), PhD or Masters theses (e.g. P. Thompson 2000, Charles 2003, Maroko 2013, Samraj 2013) or undergraduate final year projects (e.g. Hyland, 2012). There have been some investigations of citations in other types of student writing (e.g. Cumming et al 2018), but most studies of citation in ordinary student

coursework have focussed, not on typical practices, but on the attitudes of small numbers of informants, and their use of strategies to integrate reading material, avoid plagiarism and express their own voice. In the special issue on source use in *Journal of English for Academic Purposes* (2013 vol 12 issue 2), for example, all seven articles examine students' self-reported evaluations of their practice. In contrast little has been discovered about the range of citation patterns undergraduates typically produce, across disciplines, with and without an argumentative purpose.

Of course argument is an important skill for students, and argumentative essays of various types are assigned to students in most disciplines. In order to argue successfully, student writers have to learn how to identify and evaluate the source of claims and to establish their own contribution. However it is a mistake to assume that student writing necessarily requires the positioning of the writer in relation to other competing textual voices. In their analysis of assignment genres, Author and other (2012) discuss some cases where student writers enter into dialogue with authoritative academic or professional sources, but also others which interpret non-academic sources such as novels, historical documents, medical records or economic data, and still others where no arguments are put forward at all, but events are reported, or the state of current knowledge on a given topic is presented to a new readership. Moreover, although appropriate citation practices are sometimes regarded as a useful indicator of overall student achievement (Shaw 2013), it would be dangerous to assume that all good student writing must contain citations, or that every student proposition must be accompanied by a reference to someone else's proposition, presented at some earlier point in time. Bakhtin (1986:89) argues that "Our speech, that is, all our utterances …… is filled with others' words, varying degrees of otherness or varying degrees of "our-own-ness," varying degrees of awareness and detachment". This being the case, although no words or ideas can be entirely our own (whether we know this or not), those with a high degree of 'our-own-ness' might legitimately be presented without citation. Indeed some tutors might welcome citation-free writing, if the writer has managed to internalise and reinterpret thoughts that are in current circulation, or if the writer's own empirical observations provide sufficient evidence for their claims.

Those few academic writing textbooks that delve deeper into the underlying purposes and linguistic features of citation tend to ignore the uses (and non-uses) of citation in non-argumentative academic writing. Taking their lead from the most influential researchers in applied linguistics, they often refer to the behaviour of expert writers, especially the authors

of research articles, rather than drawing on evidence from university assignment writing. Swales and Feak (2004: 251-2), for example, invite graduate students to discuss the idea that writers cite to acknowledge intellectual property rights, or as a marker of respect to prior scholars, or as "a kind of mutual reward system", or to create a research space for their own work. All of these are reasons for citation in research contexts, and may apply to student researchers at doctoral level, but are less likely to fit the purposes of students at lower levels of study. Leki (2007) and Thompson et al. (2013) both draw attention to the fact that in some disciplines of undergraduate study there is no research-based writing. Undergraduates do sometimes cite for the same purpose as research writers do, "to convince the readers that … claims are justifiable and significant" (Bloch 2010: 222), but most student genres other than the dissertation/thesis are not primarily concerned with justifying the writer's own original research. As most undergraduate writing is not read by anyone but a tutor or examiner there is no real need for writers to position themselves in relation to other members of the discourse community. Moreover, whereas new research territory is generally established by way of arguments from the 'research world', many student assignments require the extraction of facts from 'real world' sources such as industry or government documents. Sources can also serve as objects of study (for example for literary or linguistic analysis), or as "shorthand references to procedures" (Hyland 1999: 347). And students may, of course, cite simply to show their tutors what they have read – citation as a 'performance' (Harwood and Petrić, 2011).

This study sets out to identify the different features of citation in assignments from the British Academic Written English (BAWE) corpus[1], bearing in mind that different practices are appropriate in different fields and for different purposes. It aims to broaden understanding of the range of behaviours relating to source use that are expected of undergraduate students in British universities, and possibly in universities in institutions in other countries where English is used as the medium of instruction. It also presents new techniques for identifying various forms of citation in the BAWE corpus and (subject to suitable modifications) in other similarly-annotated corpora.

Given that citation is a peculiarly academic practice, and that it takes time for novice academic writers to develop the skills needed to select and incorporate sources to suit their different purposes and task requirements (see, for example, Mansourizadeh and Ahmad 2011,

[1] Website information anonymised

Thompson et al 2013), I hypothesise that students will cite more frequently and in a greater variety of ways as they progress through their years of study. As it has often been noted that citation practices vary across disciplines (see, for example, Charles 2006, Hyland 1999) and genres (see, for example Petrić and Harwood 2013, Thompson and Tribble 2001), I hypothesise that there will be considerable variation in citing styles in different disciplines and for different types of assignment, regardless of level.

2. Methodology

The BAWE corpus contains 2,761 examples of proficient written coursework, distributed more or less evenly across four levels of study (first year undergraduate to Master's level) and four disciplinary groupings: Arts and Humanities (AH), Life Sciences (LS), Physical Sciences (PS) and Social Sciences (SS). In the corpus as a whole, just over 70% of the assignments were submitted by students who self-reported as having English as their first language.

BAWE is part of speech tagged with the UCREL CLAWS tagset (Garside, 1987), Version 7. In order to map changes to citation practices in different disciplinary groupings, across years of undergraduate study, 12 subcorpora of BAWE were created, containing assignments written in each of the four domains in the first, second and final year of Bachelor degree programmes. (The final year, Level 3, represents the third year of undergraduate study, or the fourth if the student had spent the third year abroad or in employment). Undergraduate dissertations and Master's degree assignments were not included. In the undergraduate component of the corpus used for this study (2170 assignments), 1682 assignments were submitted by students who self-reported as having English as their first language, and 488 were submitted by students who self-reported as having another first language. In this group, 29 first languages were reported (with various dialects of Chinese conflated). All the assignment authors were enrolled in the UK on British university degree courses, and had been awarded high grades by their lecturers for the assignments they contributed to the corpus.

Two broad types of citation style were investigated: 'author-date' and 'number-author'. Following Swales (1990), author-date citations were treated as either integral, where the name of the cited author occurs in the citing sentence and the date (and/or page number) is provided in parentheses, or non-integral, where both the name and the date and/or page number are provided in parentheses. Number-author patterns were investigated in a more

limited way, because although the number-author system is common, patterns of use are more difficult to identify.

Searches were conducted across the three levels and the four disciplinary groupings, and within specific disciplines. All the subcorpora were created and stored in *SketchEngine*[2], the online corpus query tool **(**Kilgarriff et al. 2004, 2014), and were examined using *SketchEngine*. Frequencies presented in the Results section are those calculated by *SketchEngine*. The searches were formulated using corpus query language (CQL), elaborating on queries described in Author (2014), and improved in line with suggestions made by Ian Johnson, whose Masters dissertation compared citations in BAWE with those in a corpus of Chinese student writing (Johnson 2015). Student writing does not adhere to formatting and punctuation conventions to the same extent as published writing, so allowances were made for variation in this respect, and for the optional inclusion of additional words within citation sequences.

A basic search pattern for integral, 'author prominent' citations would take the form of 'name (date) verb'. There are 1,929 cases (231.4 per million words) of this pattern in the BAWE corpus, for example:

1. As Archer (2001) says, 'science is located in the social world'. (0190b LS level 2)
2. Oliver (1980) referred to contentment as a state. (6111d LS level 4)
3. Howe (1998) cites the work of Murry. (0082d LS level 2)

However a simple search like this also identifies 'name (date) verb' patterns within numbered references in footnotes or at the end of assignments, while failing to identify the many variants of in-text citation that occur within the body of assignments. The following integral citations, for example, would not be detected by a simple 'name (date) verb' search because other word forms occur before, within or after the parentheses:

4. Schapiro et al. (2001) demonstrated (3060a SS level 4)
5. Lipton (1991,p.419) refers (0177b AH level 3)
6. In 1991, Aram et al (Menyuk, 1995:4) reviewed (6206b AH level 3)
7. Batliwala (1993 cited in Sen, 1997: 2) brings into focus (0422a SS level 4)
8. Piaget (as cited in Rubin, LeMare and Lollis, 1990) offers (0014c LS level 2)

[2] http://www.sketchengine.co.uk

9. Gisela Bock (in Bridenthal et al 1984) argues (0408f SS level 3)
10. quoted by Aitchison (1989:260) as claiming (6055a AH level 1)
11. supported by Kjelsan et al (2004) who found (0014a LS level 2)
12. In Brown et al (1994) it says to look for (3101b SS level 1)
13. In another study by Kerr and MacCoun (1985) they demonstrated (0014e LS level 3)
14. Lloyd and Lishman (1975) originally demonstrated (0016a LS level 1)

For the purposes of this study the simple 'name (date) verb' query was therefore refined by allowing the optional inclusion of 'et al' (with or without a full stop) after the proper noun, up to five additional words on either side of a bracketed four digit number, and an optional additional word before the verb. This query (Query 1) was also designed to exclude 'name (date) verb' patterns that crossed sentence boundaries, and citations within bibliographies, footnotes and endnotes. A modification of Query 1 was created for references within brackets to a page, verse or line in an external source with no mention of a publication date. This query had to be manually filtered very extensively, however, as the page, verse or line was not always mentioned explicitly, and numbers within brackets do not necessarily indicate a citation.

The full form of Query 1 is listed in the Appendix, together with subsequent queries. These identified further adjunct and passive patterns in integral citations (Queries 2 and 3), non-integral citations (Query 4), citations using number-author systems (Query 5), and the use of *op. cit.* and *ibid.* (Query 6).

Query 1 found many examples of the 'name (date) verb' citation pattern, and was a useful first step towards identifying the reporting verbs used in integral citations. It identified some cases where the 'reporting signal' (G. Thompson 1994) was carried by an adjunct rather than a clause, as in Examples 15 - 18:

15. According to Foundation of Engineering, Holtzapple Reece (2003) torque is a twisting force. (0008b PS level 1)
16. According to Alston (1880) and Watt (1905) it arrived in the West of Scotland in the early 19th century. (6011f LS level 3)
17. According to Bowie and Buttle (2004) to satisfy said needs and wants, the guests have to be the key element. (3022a SS level 3)

18. Figure 1, obtained from Balchin et al (1995), shows that if subsidised, bus operators could cut fares. (3095e PS level 1)

However, although Query 1 found patterns where verbs closely followed the bracketed dates as in Examples 15-18, it did not identify patterns where the verb was at a greater distance, as in "According to Segal (1978), youngest children are unlikely to be successful". Query 2 was created to find these patterns. It searched for sequences beginning with 'in', 'as', 'from' or 'to', with an optional possessive 's' after the proper noun, and an optional comma immediately after the bracketed date, followed by a noun, adjective, preposition or determiner rather than a verb.

Although some integral author-date citations containing long passive constructions with 'by' were picked up by Queries 1 and 2, for example "Some predictions made by Robertson et al (1999) make this even clearer" (Query 1), and "As suggested by Oakley (1981), motherhood is an essential stage of development" (Query 2), in order to capture as many integral citations containing long passives as possible, a third query (Query 3) was created. This query specified that a verb followed by 'by' should occur before the proper noun and bracketed date.

As far as possible, stray results from Queries 1-3 that were non-integral rather than integral citations, as in "research undertaken by Wilkinson (Giddens 2001)" or "the characters in Emma (Austen, 1816)", were removed using the *Sketch Engine* 'filter' facility. Results with dates which did not refer to a year of publication were also removed, as in Examples 19 - 22:

19. Flourens (1794 - 1867) studied Gall's work on phrenology. (0031c LS level 1)
20. Mary Quant (born February 11, 1934) is an English fashion designer. (0227d PS level 1)
21. The more successful English settlement of Virginia (from 1607) was threatened by local tribes. (0280a AH level 1)
22. Significant income was generated by Baxi (£11.8m in 2003, £ 7.2m in 2002). (0027a PS level 3)

For the non-integral, 'information prominent' '(name + date)' citation pattern a fourth query was created (Query 4). This query looked within sentences for brackets containing proper names and up to five additional words on either side of a four digit number. As with Queries 1 to 3, examples from bibliographies, footnotes and endnotes were excluded. Stray results

where place names or course names rather than author names were placed in the parentheses were filtered out. Examples 23-25 were identified using Query 4:

23. ideas of nationalism often associated with both the ideology and actions of racism (Fenton, 1999) (0001a SS level 1)
24. the coping technique adopted to achieve what is believed to be desirable (Gordon, 1990) (0014a LS level 2)
25. his theory on blackbody radiation (Planck 1900) for which he was later awarded a Nobel Prize (0109a PS level 1)

Additionally, a modified search for a numeral of up to three digits within brackets was used to find any instances (in either integral or non-integral citations) of Modern Language Association (MLA) author-page citations, where the reader is referred to a page, verse or line in the source under discussion, rather than a publication date. Results from searches for in-text 'author-page' citations had to be extensively filtered, manually, however, as numbers are used not only to refer to dates, pages or parts of texts, but also (more commonly) to refer to numbered reference lists, visual elements and examples.

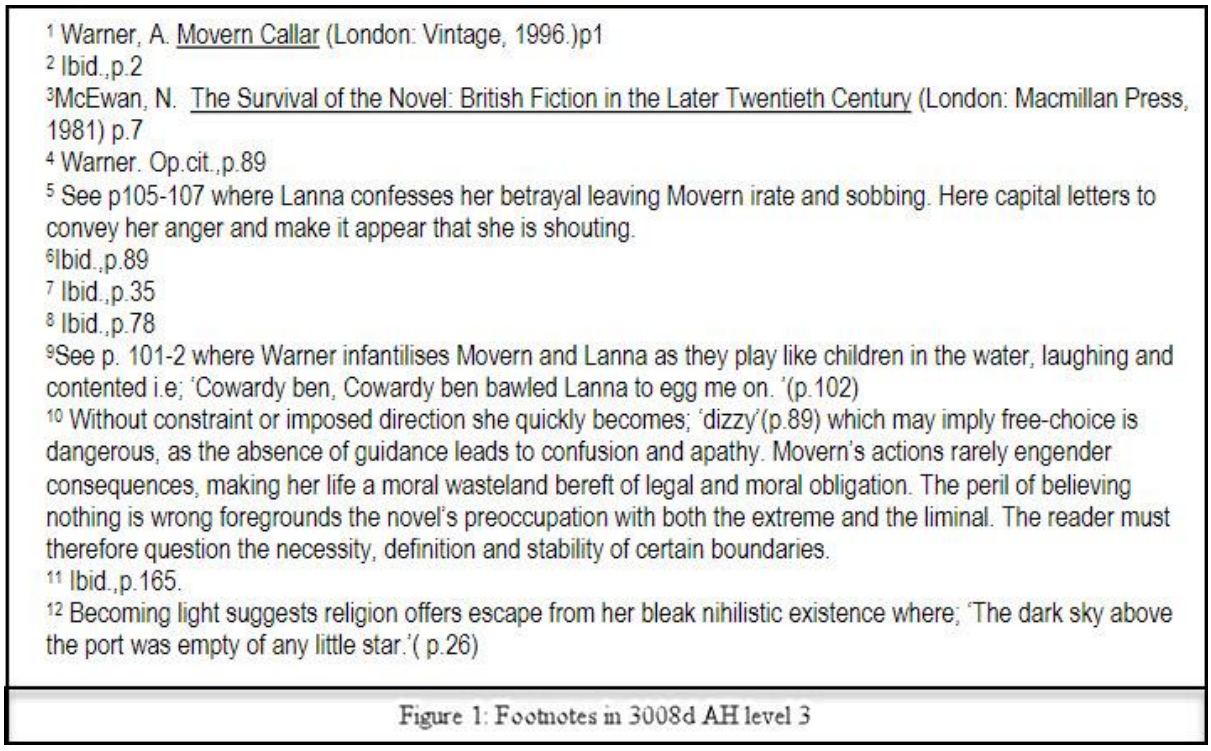
Query 5 was designed to investigate the use of number-author systems, in which a number, in brackets or as superscript, is given to each cited source in the body of the text, and fuller details of the source are provided in footnotes or endnotes. Searching for number-author citations is not straightforward because numbers within assignments do not necessarily refer to cited sources, and when using a number-author system a reference to a proper name or a publication year is not necessarily integrated in the body of the text. Moreover when citations are included in numbered footnotes, the bibliographical details may be divided between the note and the body of the text, the note containing the title of the source instead of, or as well as, the author's name.

Because of this wide variety of numbered citation formats in the BAWE corpus, Query 5 could not specify exact search patterns, but took the more broad-brush approach of searching for paragraph-initial one or two digit numbers (optionally within round or square brackets) within notes and bibliography sections. This query provides information about the amount of use of numbering systems in the corpus, although it does not distinguish between references to sources and other types of information that might be included in a numbered list - the "comment, explanation, or information that the text can't accommodate" (*MLA Style Centre*, undated).

Example 26 was retrieved using Query 5:

26. [1] - G.F. Knoll, "Radiation Detection and Measurement", (Wiley, New York,1979) [2] - University of Physics Department, "Multi-channel Analyser and Genie 2000 User Guide"[3] - University of Physics Department, "S3 Gamma-ray Spectroscopy" Second Year Lab Script [4] - Robert C. West, "Handbook of chemistry and physics - 49th edition". (0051c PS level 2)

Figure 1 shows footnotes in a Humanities assignment



[1] Warner, A. Movern Callar (London: Vintage, 1996.)p1
[2] Ibid.,p.2
[3]McEwan, N. The Survival of the Novel: British Fiction in the Later Twentieth Century (London: Macmillan Press, 1981) p.7
[4] Warner. Op.cit.,p.89
[5] See p105-107 where Lanna confesses her betrayal leaving Movern irate and sobbing. Here capital letters to convey her anger and make it appear that she is shouting.
[6]Ibid.,p.89
[7] Ibid.,p.35
[8] Ibid.,p.78
[9]See p. 101-2 where Warner infantilises Movern and Lanna as they play like children in the water, laughing and contented i.e; 'Cowardy ben, Cowardy ben bawled Lanna to egg me on. '(p.102)
[10] Without constraint or imposed direction she quickly becomes; 'dizzy'(p.89) which may imply free-choice is dangerous, as the absence of guidance leads to confusion and apathy. Movern's actions rarely engender consequences, making her life a moral wasteland bereft of legal and moral obligation. The peril of believing nothing is wrong foregrounds the novel's preoccupation with both the extreme and the liminal. The reader must therefore question the necessity, definition and stability of certain boundaries.
[11] Ibid.,p.165.
[12] Becoming light suggests religion offers escape from her bleak nihilistic existence where; 'The dark sky above the port was empty of any little star.'(p.26)

Figure 1: Footnotes in 3008d AH level 3

A final query (Query 6) was conducted to gauge the prevalence of the abbreviations *op. cit.* (from the Latin phrase *opere citato*, "in the work cited") and *ibid.* (from the Latin adverb *ibidem*, "in the same place"). For this query entire texts were searched, including their footnotes and endnotes.

Examples 27 and 28 were retrieved using Query 6

27. Peter Lambert, "The Professionalisation and Institutionalisation of History" in Berger et al, op cit, p. 45 and Georg G Iggers, op cit , p. 24. (0042c AH level 3)
28. These sites, possibly late Neolithic settlements (ibid), are contemporary with the stone phases of Avebury. (6171j AH level 3)

This study did not investigate 'follow-on' reporting clauses referring to previously-mentioned authors but not previously-mentioned source details (Example 29), or 'implicit attributions' (Williams 2010), where instead of precise source details writers make general reference to the work of other authors (Example 30).

29. Perhaps the connection with this aridity of the psyche develops the blood metaphor identified by Sinclair. (0024h AH level 3)
30. Recent criticism has argued that the Cold War split has caused intellectuals to make an over-simplified distinction between 'individualistic liberalism and state collectivism'. (0003j AH level 1)

Shaw (1992) and Charles (2006) were able to count follow-on references in their more manual corpus analyses, but references without precise source details were considered too varied in form to be identified systematically using corpus query methods. Manual analysis of the corpus identified some cases of implicit attribution, however, and examples are provided in the results section, to illustrate the way this technique was used by students as a means of avoiding negative evaluation of the work of prior researchers.

In the results provided below, duplications in the data have been automatically discounted, using the 'overlap' filter in *SketchEngine.* Because the subcorpora are of slightly different sizes, the results have been normalised in terms of their relative frequency in each individual subcorpus, rather than in the corpus as a whole.

3. Results

This section examines corpus output for each of the queries in turn: Query 1 ('name (date) verb' integral citations), Query 2 (the adjunct rather than clausal pattern in integral citations), Query 3 (long passives with integral citations), Query 4 (non-integral '(name + date)' citations, Query 5 (numbered citations) and Query 6 (op. cit. and ibid). Writers' stylistic choices are compared across disciplines and levels, and the possible reasons for these choices are discussed, bearing in mind that citation function is necessarily open to interpretation. Outputs for the queries concerning integral citations (Queries 1, 2 and 3) will be discussed in Subsection 3.1., and outputs for Queries 4, 5 and 6 will be considered in the Subsections 3.2., 3.3., and 3.4.

3.1. Integral citations

The results for Query 1 (‘name (date) verb’ integral citations) are presented in Table One. They show a quite dramatic rise in the use of integral citations across the levels of study, except in the Physical Sciences where this citation type occurs very rarely. (There were only 24 results for Query 1 across the entire Physical Sciences subcorpus, levels 1 to 3).

	Arts & Humanities	Life Sciences	Physical Sciences	Social Sciences
Level 1	75.13	245.01	29.05	166.73
Level 2	193.87	462.82	16.55	295.46
Level 3	287.15	717.23	15.88	526.46

Table 1: Results for Query 1, per million words

The results for Query 1 included some instances where a secondary source was acknowledged, as in Examples 31 - 37:

31. Haan writing in Leitner (1992) agrees with this method. (6062a AH level 2)
32. Coates refers to a study by Kramer (1974) which showed that students of both sexes characterised women's speech as 'stupid, vague, emotional, confused and wordy'. (1989: 65). (6067a AH level 1)
33. Hufnagel (1979) is quoted by Aitchison (1989:260) as claiming that sound misplacement occurs. (6055a AH level 1)
34. Deuchar (1988 as in Holmes, 1995: 2) suggests that those who are powerless must be polite. (6206a AH level 3)
35. 10 questions previously used by Sadler (cited by McManus, 2002) (0033a LS level 1)
36. One theory about the learning of aggression was proposed by Huesmann (as cited in Josephson, 1987). (0262e LS level 2)
37. "elastic and flexible" as described by Johnston (in Field and Leicester 2000) (0320a SS level 2)

Table Two shows filtered results for Query 1 which gauge the extent of this practice across the levels and disciplinary groups:

	Arts & Humanities	Life Sciences	Physical Sciences	Social Sciences

Level 1	3.34	36.61	0	14.23
Level 2	7.97	15.89	5.52	23.70
Level 3	13.84	8.71	0	29.89

Table 2: Query 1 citations which acknowledge secondary sources, per million words

In Arts and Humanities and the Social Sciences these references to secondary sources seem to increase slightly at later levels, whereas they decrease in the Life Sciences. They occurred most often in the more applied disciplines, particularly Applied Linguistics, Health, Psychology, and Hospitality, Leisure and Tourism Management. References to secondary sources were also found in non-integral citations (Query 4), and presumably tutors were prepared to accept their occasional use on the grounds that undergraduates would have difficulty obtaining access to certain books and journals, especially within the timescale set by coursework schedules. It is likely that many more citations in the corpus were secondary, but were not acknowledged as being so. In her qualitative study of postgraduate student writing, Pecorari (2006) was able to establish that it was common for secondary citations to remain unacknowledged. She was only able to do this, however, by comparing passages from the student texts with their sources. This would be impossible in a corpus-based study, both because of the number of citations involved, and because many sources would be difficult to obtain.

Table Three shows results for Query 2 (the adjunct rather than clausal pattern in integral citations). A slight general increase in use is indicated across the levels of study, although numbers were extremely small (85 in total). The numbers in brackets in the Physical Science column refer to the raw occurrences of the adjunct pattern in these sub-corpora, and reveal how very rare this pattern was in this disciplinary group.

	Arts & Humanities	**Life Sciences**	**Physical Sciences**	**Social Sciences**
Level 1	5.01	14.08	5.81 (2)	10.17
Level 2	9.30	13.90	2.76 (1)	22.12
Level 3	17.30	14.52	3.97 (2)	39.86

Table 3: Results for Query 2, per million words

In Query 2 results, “According to X” was by far the most frequent way of referring to sources, as in “According to Cassell (2001) this does not pose a problem”. Other similar expressions also occurred, however, underlined in Examples 38 - 42:

38. As shown in Kydland and Prescott (1977), in a dynamic two period context, rules-based policies are time-consistent and optimal (0301a SS level 2)
39. as noted by Lefever (1988) technology has not reached a point where it can totally replace human intuition. (3082b SS level 2)
40. as pointed out by Wakelin (1984) exploitation of the environments provided by the bodies of living animals is, in essence, no different from the exploitation of the environments provided by the seas (6147c LS level 1)
41. as suggested by Oakley (1981), motherhood is an essential stage of development. (0214c SS level 3)
42. In the words of König (1998), intervention was seen no longer a right or a privilege (3042h AH level 3)

The query also revealed the use of the name-date combination as a noun phrase representing an element in the source text, as in Examples 43 – 46:

43. As illustrated in Table 3, taken from Harris and Grace (1999) stranger rape were just 11% of the overall reported rapes (0214b SS level 3)
44. The rising importance of NGOs in Post-Cold War diplomacy is best illuminated in McRay (2001) by what he calls The New Diplomacy (6180b SS level 3)
45. The reader is referred to Giddens (1968) for a discussion of Parsons (0004d SS level 3)
46. Referring to Caparo v Dickman (1990), House of Lords held that the defendant must reasonably have anticipated that the advice would be acted upon (0153b SS level 1)

Alternatively the name-date combination, often preceded by a determiner, was used to specify a particular type of test, experiment, model etc., with no apostrophe needed after the author’s name. For example:

47. the MRT Test based on a Vandenberg and Kruse (1978) type test (0363d PS level 3)
48. the results from the Rice and Woodsmall (1988) experiment showed that whilst five-year-old children were very efficient at learning the meaning of words (6067e AH level 2)

49. [the] Farrar and Selwyn (1967) model represents the after-tax income of an investor (0287d SS level 3)

In such cases the source is named as a kind of shorthand reference, a typical technique in 'clipped' methods texts (Swales 2004), and one that assumes expert knowledge on the part of the reader. In contrast elaborated or 'slow' texts use descriptions rather than citations to explain the methodology adopted (Bloor 1999). This has relevance for the teaching and learning of academic writing, as EAP materials often focus on the more explicit 'slow' reporting of experimental procedure (for example through the production of process texts), and ignore less overt ways of referring to method.

The corpus output provides examples of apostrophe use that might be examined in the EAP classroom, especially when students are unsure about the majority practice in their discipline. Some apostrophe positions were found to be very rare within citations, and apostrophe use generally was less uniform in non-clausal citation patterns, where the date in brackets interrupts the flow between the producer and the product.

In some Query 2 results the apostrophe was missing, although formally required:

50. Atkinsons (1984) theory of a flexible firm (3074a SS level 3)
51. Gibbs (1988) reflective cycle (3064g LS level 2)
52. Hughes (2002) list of wandering-management guidelines (3076c LS level 3)
53. Guay and Harford (2000) research discovers the price reaction. (0287d SS level 3)

Very occasionally the apostrophe was placed after the author's name and before the date in brackets, as in Examples 55 - 58:

55. The first section contained 35 items selected from Folkman and Lazarus' (1985) *Ways of Coping* Questionnaire. (0016c LS level 2)
56. I found Hughes' (2002) guidelines helpful. (3076c LS level 3)
57. to borrow Soames' (2006; 21) example (0303a AH level 2)
58. Godfrey and Clarke's (2000) and World Tourism Organization's (1994) approach to tourism planning (3153a SS level 3)

However there was only one instance in the entire corpus of the possessive 's' after the date in brackets:

59. Hoque (1999) and McGunnigle (2000)'s findings are focused on the hotel sector. (3018c SS level 2)

Perhaps this pattern was disliked by the student contributors because it sounds clumsy and does not usually appear in academic reading materials.

Table Four shows the amount of long passives with integral citations in the corpus, as identified by Query 3. The numbers of instances of this pattern were relatively small (there were only nine in the whole of the Physical Science subcorpus, levels 1-3), but the results once again indicate a general tendency for the use of this pattern to increase across the levels.

Query 3	**Arts & Humanities**	**Life Sciences**	**Physical Sciences**	**Social Sciences**
Level 1	5.01	22.53	14.53	12.20
Level 2	21.25	25.82	8.28	34.76
Level 3	19.92	43.56	1.98	39.86

Table 4: Results for Query 3, per million words

In order to investigate the number and type of finite active lexical verbs occurring in the 'name (date) verb' pattern, the results for Query 1 were returned to, and interrogated further. Although 185 verb types (lemmas) were identified, many of these occurred only once or twice, for example *collate, echo, retaliate* and *unearth*. The numbers of different types at each level and in each domain are presented in Table Five. A steady rise in the variety of types was noted in all domains except the Physical Sciences, with Level 3 students using about twice as many different reporting verbs as Level 1 students. This reflects the increasing lexical richness of student writing as they advance through the levels of study. The apparent lack of development amongst Physical Science students in this respect is probably simply due to the fact that they do not typically use the 'name (date) verb' pattern.

	Arts & Humanities	**Life Sciences**	**Physical Sciences**	**Social Sciences**
Level 1	22	38	6	34
Level 2	49	61	5	59
Level 3	58	71	5	63

Table 5: Active finite lexical verb types in 'name (date) verb' clauses

Verb choices in citations position the writer metadiscursively in relation to the source text. The most frequent lemmas and word forms are shown in Table Six, with a tentative categorisation in terms of the activity types identified by Hyland (1999).

Lemma	**No.**	**Activity Type**	**Verb form**	**No.**	**Activity Type**
suggest	147	Discourse	found	82	Research
argue	105	Discourse	suggests	71	Discourse
find	88	Research	argues	61	Discourse
state	83	Discourse	suggest	46	Discourse
believe	50	Cognition	states	41	Discourse
cite	49	Discourse	cited	41	Discourse
explain	41	Discourse	points (out)	36	Discourse
point (out)	40	Discourse	suggested	30	Discourse
describe	37	Discourse	argue	28	Discourse
claim	32	Discourse	state	27	Discourse
define	29	Discourse	describes	23	Discourse
show	27	Research	believe	21	Cognition
propose	26	Discourse	believes	20	Cognition
conclude	23	Cognition	showed	19	Research
highlight	23	Discourse	claims	19	Discourse

Table 6: Frequencies and functions of the top lemmas and verb forms occurring in ‘name (date) verb’ clauses

Most of the most frequent reporting verbs were in the ‘discourse’ category, involving verbal expression, although *found,* in the ‘research’ category, was the commonest form. These verb lists are derived from Query 1 results, so they show the number of uses of the verbs after a date (in parentheses). In these contexts, *cite* is used like other reporting verbs, rather than to signal a secondary source. It is sometimes followed by *that*, as are most of the other reporting verbs on the list. *Cite that* may sound odd, or stylistically clumsy, to some readers; it occurs six times in the undergraduate corpus. The slightly less problematic form *highlight that* occurs 70 times, as in Examples 60 - 62:

60. Kotas (1996) cites that the fast food restaurants, McDonald's and Burger King are based on strictly standardised rules. (3152a SS level 1)

61. Kennedy (2001) highlighted that there are deficiencies in paediatric health care settings (3031b LS level 2)
62. Lewis and Chamber (2000) highlight that it is essential to not only look at the physical feature of the competition but also look at the service they provide (3022a SS level 3)

The writers' choices of reporting verbs suggest that they were generally accepting of source information (e.g. *highlight, show, point out*) or neutral towards it (e.g. *argue, believe, cite, suggest*). 'Counter-factive' verbs that indicated a source writer's critical attitude towards an earlier proposition were occasionally used, as in Examples 63 - 66:

63. Eugene D. Genovese (1974) denied that enslaved women had 'male' characteristics (0318b AH level 3)
64. Feinstein et al (2003) and Shuller et al (2002) disagree that participation in leisure courses reduces the potential or improves the mental health of those suffering from depression. (0320c SS level 3)
65. Gough and Wren (1999) disputed the findings of Goodman (1973) in their own miscue analysis task (6174e AH level 3)
66. Laughlin and Jaccard (1975) have refuted social facilitation theory. (0022b LS level 2)

However verbs which signalled the writer's own issues with the source (e.g. *challenge, fail, overlook*) did not occur in the Query 1 pattern. Negative or critical evaluations of sources were expressed in the BAWE assignments, especially in the final undergraduate year when students were becoming more confident and knowledgeable about their field of study. However these tended to be expressed as 'implicit attributions' (Williams 2010), often with hedging, an agentless passive or a vague group reference. This is a safer approach for students who are still at the periphery of their disciplinary discourse community. As these cases did not contain an explicit reference to a source, they were not identified by Queries 1 to 4, and are not included in citation counts.

Examples 67-70 were retrieved by searching for author-critical verbs in the corpus and manually selecting those that implicitly indicated dissatisfaction with source material.

67. Freud has been challenged on a personal level as well as a theoretical level with claims that he is more of a showman than a psychologist (0082a LS level 1)

68. This field had been largely overlooked with the rapid rise and advancement in electronics and computers. (0249d PS level 3)
69. This instability of anthocyanins in tissue culture and in the body often tends to be overlooked. (6008c LS level 3)
70. Many commentators believe Hume's criticisms have decisively refuted design arguments of the form presented by Paley and Cleanthes. (6197i AH level 3)

It will have been noticed that for Queries 1 to 3, which revealed the incidence of different forms of integral citation, the numbers for the Arts & Humanities were lower than those in the Life Sciences and Social Sciences. This initially appeared to be a surprising finding, as previous studies such as Hyland (1999) and Maroko (2013) found greater use of citation in the soft fields than in the sciences. Maroko concluded that whereas in the sciences empirical observations counted for more than arguments, humanities and social science dissertation writers needed to cite in order to "bring on board arguments and counter-arguments from previous literature" (2013: 128).

There seem to be two main explanations for the lower numbers of integral 'name (date) verb' citations in the BAWE corpus Arts & Humanities assignments, as compared to those in the sciences. The first of these is to do with the fact that, although argument is important in the humanities, it tends to rely on evidence from a smaller number of sources. Discussion in both expert and student humanities writing often revolves around the interpretation of Exhibits, such as works of art or literature, with only limited reference to Argument sources. (It is for this reason that the SCImago journal rank metric is adjusted to allow each citation in a humanities article to count proportionally more towards the journal impact score than it would in journals from other domains.) A second explanation for the failure of Queries 1, 2 and 3 to identify many Arts & Humanities citations is the prominence of an alternative MLA-style 'author-page' referencing system in this domain. MLA-style citations refer to the page, verse or line in an external source without mentioning a publication date.

Although a few of these kind of citations were produced by individual writers in Life Sciences (Biological Sciences and Health), in Physical Sciences (Architecture and Engineering), and in Social Sciences (Politics, Sociology and Hospitality and Leisure Tourism Management), most instances occurred in the Arts & Humanities, in English, Comparative American Studies , Classics, and Philosophy. In these disciplines page

references within brackets, without the date of publication, occurred about 125.9 times per million words. Most were linked to direct quotations, as in Examples 71 - 75:

71. McGregor disagrees saying 'in the early years the synods were meeting and, probably, the Athenians were not solely responsible for the sentencing of the Naxians.' (pg 40). (6143c AH level 1)
72. Macondo is wiped off the face of the earth, 'by the wind and exiled from the memory of men.' (422). (0322b AH level 1)
73. This is always in conjunction with opposing pernicious Christian teaching; whose standards of morality he claims are "out of harmony with the actual world" (85) and "unknown to the environment" (86). (3003a AH level 3)
74. 'Right opinion, whether evanescent or constant, is a good guide while it prevails; but its genesis is largely beyond human control.' (71) (0215b AH level 1)
75. the problem arises when Descartes suggests that body and mind are distinct from each other because of the fact that he "can clearly and distinctly understand one thing apart from the another" (page 54). (0055b AH level 1)

Typically the assignments in which citations of this kind occur centre on the analysis and interpretation of key Exhibit sources. In these cases, although full bibliographical details may be provided in a bibliography, the date of publication is unlikely to be needed as it is understood to be given information already known to readers.

3.2. Non-integral citations

In every domain non-integral citation was much commoner than integral citation (identified through Queries 1, 2 and 3). This was also found to be the case in other corpus studies such as those of Samraj (2008), who examined Masters theses, and Swales (2014) who examined the work of undergraduate and graduate students. Use steadily increased across the levels of study, as shown in Table Seven.

	Arts & Humanities	**Life Sciences**	**Physical Sciences**	**Social Sciences**
Level 1	582.64	1,999.48	92.97	1,095.95
Level 2	501.94	2,451.14	24.83	1,287.68
Level 3	1,288.72	3,258.03	83.37	1,750.45

Table 7: Results for Query 4, per million words

Non-integral citation occurred with all Bizup's source types: Background, Argument, Exhibition and Method. However, although it might be argued that attribution of any kind is an act of evaluation (Hunston 1995: 134), non-integral citations are metadiscursively most useful when the writer wants to give prominence to the original source, and subordinate his or her own voice. This is because non-integral citations do not necessitate the use of a 'reporting signal' - an adjunct or a projecting clause - which might indicate the stance of the writer. The non-integral citations identified by Query 4 often reference Background sources such as reports from companies or organisations, and accompany 'bare' declaratives where all responsibility for the proposition is attributed to the source and/or the author of the source, as in examples 76 to 82:

76. phosphorus loss can occur through soil erosion (DEFRA, 2003). (6214e LS level 3)
77. one-third to half of all diabetics already have evidence of organ or tissue damage (UK Prospective Diabetes Study Group, 1991). (6012g LS level 2)
78. each individual bat can eat up to 3000 insects in one night (RSPB, 2007). (6181a LS level 3)
79. The Adoption and Children Act 2002 (ACA 2002) now allows same-sex couples who are civil partners or cohabitants to adopt jointly. (0312c SS level 2)
80. Chinese and Italian foods are now the most popular foods with this age group (BBC Website, 2002). (3040a SS level 2)
81. Cowley Manor has recently been voted the 5th best hotel spa in the country (Condé Nast Traveller, Oct 2005). (3101e SS level 2)
82. Qantas purchased Impulse and began operating its routes under the QantasLink brand as a wholly-owned subsidiary (Qantas, 2004). (3042b SS level 2)

Non-integral citations do not require a 'reporting signal' carried by an adjunct (e.g. 'according to X) or a clause (e.g X argues that....), and because of this, sentences containing non-integral citations can be less grammatically complex than sentences containing integral citations. (Compare, for example "Mayo and Jarvis (1981) believed the answers to the question depend on the individual" to "The answers to the question depend on the individual (Mayo and Jarvis, 1981)"). The option not to use a reporting verb (an indication of the writer's stance) also means that non-integral citations do not require the writer to be so explicit about the relationship between the citing and the cited texts. It is therefore easier to cite in this way, although there is sometimes a danger that the purpose of the citation will be obscured,

especially when it is used to support an argument rather than a monoglossic assertion. For example, in the following claim from a 2nd year sociology essay it is unclear whether Gittins (1993) is cited because of his or her opinion about the importance of the stated influence, or because he or she believed that life after death was no longer considered a reality:

83. they are a way of reproducing ones self and ensuring ones characteristics remain, an influence becoming perhaps more important as life after death is no longer considered a reality (Gittins, 1993).

A lecturer marking a student's assignment might comment adversely on such lack of clarity, even if they are familiar with the source and recognise the writer's communicative intention.

Non-integral citations also have the potential advantage of making it easier to refer to multiple sources, and pile on the evidence for extensive reading without expending too many words. In practice, however, although it was fairly common to cite a secondary source alongside the original (accompanied by words such as "cited in" or 'quoted in") in the entire undergraduate corpus there were only 175 instances of writers citing more than one primary source at a time. Very few of these citations included more than two sources, and none more than three.

3.3 Numbered citations

Citations using in-text numbering to refer to sources were investigated using Query 5, which searched in footnotes, endnotes and bibliographies rather than the body of assignments. This seemed the best way to attempt to identify references using the Vancouver number-author system, an accepted method of citation in the sciences, and notes numbered in the manner authorised by the *MLA Handbook for Writers of Research Papers* (Gibaldi 2003) and the *MHRA Style Guide* (Richardson et al. 2013). It also aimed to identify other types of numbering systems used within some disciplines; the Diels-Kranz System for citing pre-Socratic philosophers, for example, and the *Oxford Standard for Citation of Legal Authorities* (OSCOLA, Meredith & Endicott 2006) for law reports and judgements.

Numbered citations can be formatted and positioned in very varied ways within student assignments, and it is probable that not all instances were identified by Query 5. The results in Table Eight reflect a greater use of numbering in the sciences, increasingly in later years of study, but, given the low numbers recorded for Physical Sciences in response to all types of query, we must conclude that that students in the Physical Sciences make very little use of

citation in any form. This is probably because knowledge claims in the writing of undergraduates in the Physical Sciences tend to be based either on received disciplinary wisdom or on information derived from the students' own calculations or experimental data. In neither case would there be a source to cite.

	Arts & Humanities	**Life Sciences**	**Physical Sciences**	**Social**
Level 1	40.07	0	142.36	0
Level 2	1.33	47.67	195.88	1.58
Level 3	3.46	133.57	293.77	13.29

Table 8: Results for Query 5, per million words

3.4. Ibid and op.cit.

Finally, the results for Query 6 (Table Nine) establish conclusively that *ibid* and *op.cit.* are a feature of the soft rather than the hard disciplines. *Op. cit.* was used relatively rarely, by only two students in Arts & Humanities, and seven students in Social Sciences. *Ibid* was more common, especially in the final level of undergraduate study.

	Arts & Humanities		**Life Sciences**		**Physical Sciences**		**Social Sciences**	
	ibid	*op. cit.*	*ibid*	*op. cit.*	*ibid*	*op. cit.*	*ibid*	*op. cit.*
Level 1	265.44	0	0	0	0	0	197.23	12.2
Level 2	241.67	2.66	0	0	8.28	0	199.08	107.44
Level 3	422.08	53.62	58.08	0	3.97	0	325.51	11.63

Table 9: Results for Query 6, per million words

4.Pedagogical implications

The undergraduate students who contributed to the BAWE corpus were all successful writers, and their assignments had been given good grades by their subject lecturers. They cited sources to achieve their own communicative purposes (to argue, explain, indicate method and display knowledge) rather than to prove their mastery of citation skills – their assignments were assessed by subject experts rather than academic writing or study skills teachers, after all.

The findings from this study are not intended to show 'correct' ways of citing sources – for this, educators and students can turn to style guides and manuals. Instead they are intended to help educators tailor their advice to students according to the writing tasks they have been assigned and the roles they wish their sources to play. The queries listed in the Appendix will all work with BAWE in *SketchEngine*, an open-access, non-subscription resource. Teachers can thus extract further, and longer, examples of citations from the corpus, and select examples from specific disciplines, genre families and levels of study. Teachers might also encourage students to investigate the results from the six queries for themselves, following the pedagogical technique of Data Driven Learning, originally proposed by Johns (1991a, b). *SketchEngine* allows links to concordance pages to be saved and retrieved; if these links are made available to students they do not need to key in queries, and can examine the concordance output interactively, in their own time.

The findings provide evidence that citation increases across levels in all disciplinary areas apart from the Physical Sciences; we might thus assume that students at higher levels of study are generally expected to cite more, and that learning about citation should be a priority for EAP students in most disciplines. EAP materials, style guides and manuals do not usually provide a wide range of examples to illustrate how citations can be worded; in contrast the BAWE corpus queries provided in this study generate references to sources expressed in a wide variety of different ways. Using these queries, teachers and/or learners can identify forms that are most common in particular disciplines, at particular levels of study, and in association with particular types of source. Noun phrases in citation patterns, for example, can be seen to refer not only to people, but also to texts, statutes, cases, models or experimental methods, while the range of reporting verbs in use by undergraduates can be seen to extend well beyond the number commonly taught in writing classes. One possible learning activity would be to place some of these verbs within metadiscoursal categories such as those developed by Hyland (1999 and 2005), in terms of the extent to which they indicate acceptance or rejection of a source proposition. Another possible activity would be for students to examine the citations in their own academic writing with reference to the findings reported in this study, and discuss the possible communicative effects of altering their wording and citation styles.

It should be noted that the study did not examine assignments where citation did not occur, although there are a considerable number of these in the BAWE corpus. For example there is a lack of citation in some Methodology Recounts and Design Specifications in the Physical

Sciences: assignments of this type can be compared to the central sections of research articles where experimental methods and results are reported, as opposed to the opening and concluding sections of research articles where contextualisation within the broader field of knowledge takes place. There is also a lack of citation in some public engagement or workplace genres in the BAWE corpus (for example some Case Studies and some pieces of Empathy Writing) where students address an imaginary audience of non-experts: in these sort of assignments students have to choose whether to write like a practitioner (who would be unlikely to cite sources) or as a student (who wants to signal knowledge of these sources to the actual academic reader). Again, educators can give more appropriate tailored advice to student writers if they are aware that citation may not be required, and that references to sources do not automatically lead to higher grades on every assignment task..

As students utilize citation in so many different ways, this is an area that is ripe for further and deeper consideration of what sort of sources should be cited, when, why, and how**.** This study set out to broaden understanding of the various ways in which undergraduates at British universities cite sources, but it did not compare the information and claims provided in citations with the content of the sources cited, and it did not attempt to discover what the writers' thought about their reasons for choosing specific citation patterns and wording. It also did not look in any detail at the text surrounding each citation, which could also have shed light on citation purposes and practices. All of these aspects of citation could be investigated more deeply using a smaller dataset, perhaps specific to a particular discipline and level.

References

Author 2014.

Author & other 2019.

Author & other 2012.

Bakhtin, M. M. 1986. *Speech Genres and Other Late Essays.* Translated by Vern W. McGee. University of Texas Press, Austin, Tx.

Bakhtin, M. M. 1981. *The Dialogic Imagination: Four Essays.* Translated by Michael Holquist, Caryl Emerson. University of Texas Press, Austin, Tx

Bizup, J. 2008. BEAM: A rhetorical vocabulary for teaching research-based writing. *Rhetoric Review* 27, 72-86

Bloch, J. 2010. A concordance-based study of the use of reporting verbs as rhetorical devices in academic papers. *Journal of Writing Research* 2, 219-24.

Bloor, M. 1999. Variation in the methods sections of research articles across disciplines: the case of fast and slow texts. In: P. Thompson (Ed.). *Issues in EAP Writing Research and Instruction.* University of Reading, CALS, Reading, UK, pp. 84-106.

Charles, M. 2006. Phraseological patterns in reporting clauses used in citation: a corpus-based study of theses in two disciplines. *English for Specific Purposes* 25, 310-331.

Cumming, A., Yang, L., Qiu, C., Zhang, L., Ji, X., Wang, J., Wang, Y., Zhan, J., Zhang, F., Xu, C., Cao, R., Yu, L., Chui, M., Liu, M., Cao, M. & Lai, C. 2018. Students' practices and abilities for writing from sources in English at universities in China. *Journal of Second Language Writing*, 39, 1-15.

Garside, R. 1987. The CLAWS word-tagging system. In: R. Garside, G. Leech and G. Sampson (Eds.). *The Computational Analysis of English: A Corpus-based Approach*. Longman, London.

Gibaldi, J. 2003. *MLA Handbook for Writers of Research Papers* (6th ed.). Modern Language Association of America, New York.

Halliday, M. A. K. 1994. *An Introduction to Functional Grammar*. London: Edward Arnold.

Harwood, N. 2008. Citers' use of citees' names: Findings from a qualitative interview-based study. *Journal of the American Society for Information Science and Technology* 59, 1007-1011.

Harwood, N. 2009. An interview-based study of the functions of citations in academic writing across two disciplines. *Journal of Pragmatics*, 41, 497-518.

Harwood, N. & Petric, B. 2011. Performance in the citing behavior of two student writers. *Written Communication* 29/1, 55-103.

Hawes, T. & Thomas, S. 1997. Tense choices in citations. *Research in the Teaching of English* 31, 393-414.

Hu, G., & Wang, G. 2014. Disciplinary and ethnolinguistic influences on citation in research articles. *Journal of English for Academic Purposes*, 14/1, 14-28.

Hyland, K. 1999. Academic attribution: citation and the construction of disciplinary knowledge. *Applied Linguistics* 20, 341-367.

Hyland K. 2012. Undergraduate understandings: stance and voice in final year reports. In: Hyland K., Guinda C.S. (Eds.). *Stance and Voice in Written Academic Genres*. Palgrave Macmillan, London pp. 134-150.

Hyland, K. 2002. Activity and evaluation: reporting practices in academic writing. In: Flowerdew, J. (Ed.). *Academic Discourse.* Longman, Harlow, pp. 115-130.

Hyland, K. & Tse, P. 2004. Metadiscourse in academic writing: A reappraisal. *Applied Linguistics* 25, 156–177.

Hunston, S. 1995. A corpus study of some English verbs of attribution. *Functions of Language* 2, 133-158.

Johns, T. 1991a. Should you be persuaded: two examples of data-driven learning. *English Language Research Journal* 4: 1-16.

Johns, T. 1991b. From printout to handout: grammar and vocabulary teaching in the context of data-driven learning. *English Language Research Journal* 4, 27-45.

Johnson, I. P. 2015. *"He said ... she said": A corpus-based comparison of the citation skills of novice Chinese L1 undergraduate students in U.K. Higher Education with those of proficient exemplars*. Unpublished M.A dissertation, School of Languages and Area Studies, University of Portsmouth, Portsmouth.

Kilgarriff, A. et al. 2014. The Sketch Engine: Ten years on. *Lexicography* 1, 7–36.

Kilgarriff, A. et al. 2004. The Sketch Engine. In *Proc EURALEX 2004*, Lorient, France; pp. 105–116.

Leki, I. 2007. *Undergraduates in a Second Language: Challenges and Complexities of Academic Literacy Development*. Lawrence Erlbaum, New York.

Marggraf-Turley, R. 2016. *Writing Essays: A Guide for Students in English and the Humanities.* Routledge, London.

Martin, J. R., & White, P. R. R. 2005.*The Language of Evaluation: Appraisal in English*. Palgrave Macmillan, Basingstoke.

Mansourizadeh, K., & Ahmad, U. K. 2011. Citation practices among non-native expert and novice scientific writers. *Journal of English for Academic Purposes* 10, 152–161.

Maroko, G.M. 2013. Citation practices in selected science and humanities dissertations: implications for teaching. *American Journal of Educational Research* 1, 126-136

Meredith, S. & Endicott, T. 2006. *OSCOLA (The Oxford Standard for Citation of Legal Authorities)* Faculty of Law, University of Oxford, Oxford, UK.

MLA Style Centre (undated) *Are notes compatible with MLA style?* Available online from Last accessed 20/05/20.

Pecorari, D. 2006. Visible and occluded citation features in postgraduate second-language writing. *English for Specific Purposes*, 25, 4-29

Petrić, B., & Harwood, N. 2013. Task requirements, task representation, and self-reported citation functions: An exploratory study of a successful L2 student's writing. *Journal of English for Academic Purposes* 12, 110-124.

Purdue Online Writing Lab (undated). *Research and Citation Resources*: *Preventing Plagiarism*. Available online from https://owl.purdue.edu/owl/teacher_and_tutor_resources/preventing_plagiarism/index.html Last accessed 20/05/20.

Richardson, B et al. 2013. *MHRA Style Guide: A Handbook for Authors, Editors, and Writers of Theses*. Modern Humanities Research Association, London.

Samraj, B. 2013. Form and function of citations in discussion sections of master's theses and research articles. *Journal of English for Academic Purposes* 12, 299–310.

Samraj, B. 2008. A discourse analysis of master's theses across disciplines with a focus on introductions. *Journal of English for Academic Purposes* 7, 55-67.

Shaw, P. 2013. Source use in academic writing: An introduction to the special issue. *Journal of English for Academic Purposes* 12, A1-A3

Swales, J. M. 2014. 2014. Variation in citational practice in a corpus of student biology papers: From parenthetical plonking to intertextual storytelling. *Written Communication* 31, 118–141.

Swales, J. M. 2004. *Research Genres: Expectations and Applications*. Cambridge University Press, Cambridge.

Swales, J. M. 1990. *Genre analysis: English in Academic and Research Settings*. Cambridge University Press, Cambridge.

Swales, J. M. & Feak, C. 2004. *Academic Writing for Graduate Students.* The University of Michigan Press, Ann Arbor, Michigan.

Thompson, G. 1994. *Collins COBUILD English Guides 5: Reporting*. HarperCollins, London.

Thompson, G. & Ye, Y. 1991. Evaluation in the reporting verbs used in academic papers. *Applied Linguistics* 12, 365-382.

Thompson, P. 2000. Citation practices in PhD theses. In: Burnard, L. & McEnery, T. (Eds.). *Rethinking Language Pedagogy from a Corpus Perspective*. Peter Lang, Frankfurt.

Thompson, P. & Tribble, C. 2001. Looking at citations: Using corpora in English for Academic Purposes. *Language Learning & Technology* 5, 91-105.

Thompson, C., Morton, J. & Storch, N. 2013. Where from, who, why and how? A study of the use of sources by first year L2 university students. *Journal of English for Academic Purposes* 12, 99-109.

Vande Kopple, W. J. 2012. The importance of studying metadiscourse. Applied Research in English, 1(2), 37-44.

Vande Kopple, W. J. (2002). Metadiscourse, discourse, and issues in composition and rhetoric. In E. Barton & G. Stygall (Eds.). *Discourse Studies in Composition*. Cresskill, New Jersey: Hampton Press pp. 91-113.

Williams, H. 2010. Implicit attribution. *Journal of Pragmatics* 42, 617-636.

Xu, F. 2016. The development features of citation competence in L2 academic writing: cross-sectional and longitudinal evidence. *Journal of Foreign Languages* 39, 73-82.

Appendix

Query 1: name (+ date within parentheses) + verb

[tag = "NP.?"] [word = "et"]? [word = "\."]? [word = "al"]? [word = "\."]? [word = "\("] [] {0,5} [tag = "MC" &word = "...."] [] {0,5} [word = "\)"][] {0,1} [tag != "VB..?|VD.|VH.|VM.?" & tag = "V..?" & textpart != "bibliography|note"] within <s/>

Query 2: name (+ date within parentheses) + noun, adjective, preposition or determiner

[word = "in|as|from|to|on"] [] {0,3} [tag = "NP.?"] [word = "\."]? [word = "al"]? [word = "\."]? [word = "\'"]? [tag = "ZZ1"]? [word = "\("] [] {0,5} [tag = "MC" &word = "...."] [] {0,5} [word = "\)"] [word = "\,"]? [tag = "NN..?|JJ.?|I.|DD.?" & tag != "VV..?"] [textpart != "bibliography|note"] within <s/>

Query 3: long passive with name (+ date within parentheses)

[tag = "VVN"] [word = "by"] [tag = "NP.?"] [word = "\."]? [word = "al"]? [word = "\."]? [word = "\("] [] {0,5} [tag = "MC" &word = "...."] [] {0,5} [word = "\)" & textpart != "bibliography|note"] within <s/>

Query 4: name + date within parentheses

[word = "\("] [tag = "NP.?"] [] {0,5} [tag = "MC" &word = "...."] [] {0,5} [word = "\)" & textpart != "bibliography|note"] within <s/>

Query 5: numbered references in bibliographies or notes

<p> [word = "\(|\["]? [tag = "M.*"] [word = "\)|\]"]? [textpart = "bibliography|note"]

Query 6a: simple search ibid

ibid

Query 6b: op. cit

[word ="op"][word = "\."]?[word ="cit"]