Research Article

Understanding Attainment Disparity: The Case for a Corpus-Driven Analysis of the Language used in Written Feedback Information to Students of Different Backgrounds

Siân Alsop, *Coventry University*

Sheena Gardner, *Coventry University*

Structured Abstract

- **Background**: Disparity of attainment between different groups of students in UK higher education has been correlated with ethnicity (UUK & NUS, 2019). For example, students who declared their ethnicity as Black were 20% less likely to graduate with a top classification than those who declared their ethnicity as White (OfS, 2018a). The causes of such attainment gaps are complex, and one important factor may be the nature of the feedback given by academic staff on assignments written by different groups of students. This paper aims to explore the feasibility of investigating this hypothesis by analyzing written feedback and looking for patterns in feedback given to different groups of students.
- **Literature Review:** Research on attainment among Black, Asian and minority ethnic (BAME) students in the UK has explored a number of aspects, and has generally concluded that there are issues of "belonging" (Richardson, 2015), particularly in institutions where the majority of academic staff and students are White, but that no single variable can explain the disparity. The wording of feedback on lower-scoring papers has been shown to be more impersonal and distant than that given to students on higher-scoring papers (e.g., Gardner, 2004), which has the

DOI: 10.37514/JWA-J.2019.3.1.04

(unintended) result of increasing the sense of belonging of higher performing students in ways that can build incrementally over the years of a degree course. While there have been many such small-scale studies of written feedback, none have aimed to collect large quantities of authentic written feedback for analysis.

- **Research Questions:** The hypotheses that drive our exploration are that written feedback information (WFI; Boud & Malloy, 2013) is worded differently to different groups of students, and that there is a direct relationship between this aspect of feedback and academic attainment as measured by grades on summative assessments. Specifically, we asked:
    1. Can a framework of WFI functions be developed for our data that share a meaningful set of attributes?
    2. Can these categories be used to differentiate WFI to different groups of students?
- **Methodology**: A small pilot corpus was compiled from written feedback comments on twelve student assignments from two large Faculties. Metadata was added to each file, and the WFI comments were annotated and analyzed according to a framework developed in a branching format through a recursive construction process informed by the literature reviewed and the data in the corpus. This technique was used to characterize the WFI styles of the two Faculties.
- **Results**: The results show that all WFI comments could be classified using the novel systematic framework developed, and that its binary nature enabled ready cross-tabulation with metadata variables. *Praise* and *critique* were found to be most frequent, with specific praise of ideas (P1A) accounting for 68% of all *praise*, and specific critique of content (C1A) accounting for 49% of all *critique*. Observations tend to be the longest feedback comments (average 15.4 words). When the two Faculties are compared, two different feedback styles are evident, with Fac1 providing more *advice*, *query*, and *observation* style feedback than Fac2, and Fac2 providing more *praise* and *critique* than Fac1.
- **Discussion**: Although the data was insufficient to focus on ethnicity, intersectionality, and attainment, it is possible to differentiate the nature of the feedback—including engagement—based on contextual variables.
- **Conclusions**: The positive results show that the framework (RQ1) and the methodology (RQ2) have been successfully developed for this limited dataset and merit application in the analysis of a larger data set to explore the complex relationships between WFI, ethnicity, and attainment.

Keywords: attainment, corpus linguistics, disparity, equity, feedback, higher education, writing analytics, written feedback information

1.0 Background

Disparity of attainment among different groups of students based on ethnicity is systemic across UK Higher Education (HE), with inherent institutional, individual, and social costs. For example, in the academic year 2016–17, students who declared their ethnicity as Black were up to 20% less likely to gain a first class (i.e., 70% and above) or upper-second class (i.e., 60-69%) degree classification compared to students who declared their ethnicity as White (OfS 2018a). This 20% attainment gap indicates little movement from the 25% gap between the same student groups recorded for the academic year 2001–02 (Connor et al., 2004, p. xiv). The "uncomfortable" societal-level disparities revealed by the Race Disparity Audit (Cabinet Office, 2017) frame an attainment gap in HE that has been described as one of the "racial injustices" that "the whole sector has shied away from accepting and tackling head-on" (Amatey Doku, Vice-President HE of the National Union of Students (NUS)). The NUS and Universities UK have recently reiterated the need for increased understanding of the causes of the attainment disparity, associating it with negative impact on sense of belonging (UUK & NUS, 2019, p. 176).

Similar patterns have been observed in other contexts. For example, Poe, Nastal, and Elliot (2019) cite US Department of Education data from 2010 that shows a 40% graduation rate among Black students in 4-year institutions compared to a 64% graduation rate for White students. Although the educational contexts are very different, this disparity is alarmingly similar to that observed by Connor et al. (1996, cited in Richardson, 2015, p. 280) who found that 65% of White students graduating from four UK HE institutions had obtained good degrees, where only 39% of non-White students had obtained good degrees. Although the US context is very different from the UK, the differences in attainment across groups of students continue to follow a similar pattern. The issue remains stubborn and poorly understood, despite ongoing calls for action.

Currently in the UK, there is a focus on the Black, Asian and minority ethnic (BAME) student population. Reference to White and BAME student populations describes collective groups based on the 21 ethnicity fields required from all institutions by the Higher Education Statistics Agency in 2018–19 (HESA, 2018). We have adopted this distinction as a starting point from which to explore our hypothesis because HESA provides a unique source of metadata relating to students' self-identified ethnicity. This said, we acknowledge the problematic nature of grouping (and so obscuring) individual ethnicities in such binary categories. We also fully expect to pay due attention to research into the intersectional operation of disadvantage on ethnic attainment disparity (Gillborn, 2015).

The recognition that the playing field is not level for BAME and White students (Richardson, 2015; UUK & NUS, 2019) poses a problem for all universities, and for wider society. Research has increasingly dismissed individual deficit theory (e.g., Mavelli, 2014), and the spotlight for understanding this disparity now rests on the institutional practices and policies that disadvantage, and so disengage, groups of students with particular sociodemographic characteristics.

The causes of such attainment gaps are complex, and one important factor may be the nature of the feedback given by academic staff on assignments written by different groups of students. This paper explores this hypothesis by developing a novel feedback framework that

is appropriate for analyzing written feedback and for cross-tabulation with metadata to reveal patterns in feedback given to different groups of students.

2.0 Literature Review

A key driver of attainment, and related engagement, is feedback, which can be defined as "a process through which learners make sense of information from various sources and use it to enhance their work or learning strategies" (Carless & Boud, 2018, p. 1). One of these sources is the written feedback information (WFI; Boud & Malloy, 2013) delivered in response to coursework assignments, which remains the primary individualized interface between academic staff and students.

Despite dissatisfaction with current practices amongst students (e.g., National Student Survey results 2013–18 (OfS, 2018b)), research has focused on the form and/or function of WFI, more than on its individual delivery/reception context. Potentially useful models for profiling the function of WFI have emerged from small-scale qualitative studies. For example, Hamp-Lyons and Chen (1999) propose eight distinct functions (see Table 1), which could be applied through analysis of the WFI alone. In contrast, Hughes, Smith, and Creese (2015) propose a multi-level framework which, to apply, also requires a greater reading and understanding of the content and context of the WFI (see Table 2). While the second approach appears to be more appropriate for our project, it does raise issues of the extent to which WFI can be accurately interpreted by researchers without full access to the relevant pedagogical context and content, especially when encoding is manual and small-scale. Both models include the core functions of *praise*, *critique*, and *advice*.

Table 1

Feedback Framework Example 1 (Hamp-Lyons & Chen 1999, 211–212)

	Category	Example
Praise	Positive comments, non-controlling	Well written! Much improved.
Criticism	Negative comments or evaluations, authoritative	Contradictory sentences. Confusing.
Imperative	Comments that tell the student writer to do or change something, usually starting with a verb in the imperative form	Be specific.
Advice	Suggestive comments often in conditional mode	Maybe you could add some details here.
Closed question	Questions that either get a "yes" or "no" as answer, or else a simple one-word answer	Is this word used literally or figuratively
Open question	Questions that require more than a "yes" or "no" answer, often starting with "what," "where," "why," "who," "when," and "how"	What does this mean?
Mechanics	Comments that deal with grammar, punctuation, spelling, word choice, etc.	Although parents permission him to…
?-	No comments except a "?," usually meaning "don't understand"	??

Table 2

Feedback Framework Example 2: Hughes, Smith, & Creese (2015, p. 1085–1086)

Category			Example
Praise	P1	Direct praise	You demonstrate a good understanding of learning theory
	P2	Recognizing progress (ipsative)	This represents a considerable improvement on previous drafts
Critique	C1	Correction of errors (e.g., numerical, verbal, referencing)	[May be actual corrections of spelling or grammar written or typed onto a script, or comments about correct procedures for referencing]
	C2	Factual critiques (of content)	Grounded theory is actually a research design
	C3	Critique of approach (structure and argument)	By putting the research approach before the outline discussion, readers would have been provided with a more focussed outline of how these issues are applied to exploring the 'Accent Method' school experience and theory need to be interwoven and equal in balance
Advice	A1	Current assignment content	More could have been written about quantitative analysis
	A2	General points that refer to the current assignment	Broader reading around the topic and of studies that employ the approach would have helped create a more nuanced discussion
	A3	General points for future learning and future assignments	In your next assignment you should flag up something like this at the beginning
Clarification requests or queries	Q	Asking learners to think more deeply about their work and generate actions themselves can be achieved through questioning and dialogue	I am not clear what assumptions you refer to
Unclassified statements	O	Neutral comments, for example, that describe the piece of work but do not make any judgement, are unclassified	

Going beyond individual models of WFI, systematic literature reviews base their advice on investigation of the effect of feedback (e.g., Kluger & DeNisi, 1996). Feedback that enables students to improve their performance is considered to have a greater effect on achievement than simple praise or critique (Hattie & Timperley, 2007, p. 84; Kluger & DeNisi, 1996), but arguments for the motivating (as well as demotivating) potential of both positive and negative feedback exist, depending on delivery choices (Pit & Norton, 2017, p. 503). The direction and focus of feedback are often flagged up as important. A distinction is made, for example, between information that enables students to understand where they are going (feed up), how they are going (feed back), and where they go next (feed forward) (Hattie & Timperley, 2007, p. 87). Hattie and Timperley (2007) influentially identified four levels of feedback focus. The third level emphasizes the role of students as active participants in a dialogue about learning, which underpins a progression from feed back to feed forward to Feedback Mark 2 (Boud & Molloy 2013; Tai et al., 2018) that is receiving increasing attention.

Dependent on focus and direction, the WFI academics give to students can open up this dialogue and have an impact on related engagement (Hughes, Smith, & Creese 2015, p. 1092). There is little doubt that effective feedback is actionable; emphasis is on enabling students to use input to improve future work, or to "close the loop" (Carless & Boud, 2018). Generally, however, while these frameworks and models usefully account for the functions of WFI and understanding of pedagogic goals, they tend to obscure the individual contexts in which WFI is delivered and received. In other words, this approach to understanding WFI and its related impact assumes a consistency of student experience that is unlikely to exist.

In terms of form, corpus linguistic studies have explored the genre features of written and/or spoken academic feedback. Lee (2013), for example, examined the rhetorical move structures of and certain linguistic features within a corpus of WFI given to 100 Humanities texts. Gardner (2004) identified differences in the wording of spoken and written suggestions and advice given as feedback on high- and low-scoring texts in Applied Linguistics. Some work focuses more specifically on the language of "best feedback" (e.g., University of Edinburgh, 2017). In terms of data collection, however, such studies do not tend to extend beyond a single discipline or cohort. Such analyses of feedback in isolation fail to account for the unique circumstances of the recipient, such as their university entry route, what they study, or their self-identified ethnicity.

It follows that the unconscious biases that we know to exist within institutions (McCormick, 2015) may also have an impact on the linguistic form of WFI. Perhaps more likely is that the type of WFI that has negative impact on student motivation, such as excess or unconstructive criticism (e.g., Pit & Norton, 2007), augments the lack of belonging already experienced by certain groups of students within HE, which in turn impacts grade attained (e.g., UUK & NUS, 2019). This relationship between the type of feedback received, the sociodemographic characteristics of the recipient (such as ethnicity), and grade attained is a consideration that has not been explored in depth.

3.0 Research Questions

Our hypothesis is that a relationship exists between the written feedback information and the grade that students receive for summative assessments, and that this written feedback information is worded differently to different student groups. To test this hypothesis, we asked:

1. Can a framework of feedback functions be developed for our data that share a meaningful set of attributes?
2. Can these categories be used to differentiate feedback to different groups of students?

4.0 Research Methodology

This is a small-scale corpus study that involves building a corpus of WFI and analyzing the language of the WFI. The data is then examined to explore the frequency in which different linguistic patterns are used with different groups of students.

4.1 Corpus Data

The data for the corpus was obtained from two large Faculties[1] within a post-1992 UK Higher Education Institution[2]. The two Faculties were chosen because: a) they provide a snapshot of disciplinary difference; b) they have received different satisfaction rates in the most recent National Student Survey (OfS, 2018b), especially in questions relating to assessment and feedback; and c) their programmes are externally accredited by professional bodies and standards, which means that effective feedback is critical to the professional development of their student bodies. Although the names of the Faculties should remain anonymous for ethical reasons, we had no reason to believe that there would be inherent differences between them which might cause different satisfaction ratings. In both Faculties, over 50% of papers receive WFI electronically through the institution's virtual learning environment (VLE). To ensure compliance with ethical approval obtained, the data was anonymous, and no contact with the academic markers[3] or the students whose work was marked was possible. Similarly, although each Faculty comprises three or four Schools, and each School offers dozens of degree programmes, School and programme level information is not included here to ensure feedback comments cannot be linked to individuals giving or receiving feedback.

The dataset used to build a small corpus for the manual analysis and framework development in this paper comprises WFI delivered to 12 assessed student papers (source texts) drawn from four undergraduate modules[4] that contribute to degree classification, two

[1] Faculties in the UK HE are generally composed of Schools and/or Departments, clustered around disciplinary commonality, for example a Faculty of Arts and Humanities or a Faculty of Health and Life Sciences.

[2] Post-1992 universities tend to be former polytechnics or colleges that are city-based and attract a diverse student population.

[3] Marking is done by members of the teaching team (e.g., a full Professor or a Lecturer) under the direction of the Module Leader.

[4] Modules in this case are credit bearing, assessed units of teaching that typically span one semester and include 3–6 hours per week of classes.

modules from each Faculty. For each module, WFI in response to three source texts were selected: one "low" band (30–45), one "middle" band (46–59), and one "high" (60+) band (where a mark of 30% may be condoned, a mark of 40% may be a straight pass, marks in the 50s are good, marks in the 60s are desirable, and marks over 70 fall in the highest band). The source texts were all identified as broadly of the report genre[5] and varied in length from 1,584 to 5,687 words. The extent of the variation in raw text length presents an obstacle to accurate comparison of the occurrence of WFI functions across source texts. The variation was mediated by normalizing results per 10,000 source words, which enabled, for example, more accurate comparisons across variables such as disciplinary group, as operationalized here by the two large Faculties. It is not assumed that there is a direct relationship between the word count of a source text and the number/nature of comments given in WFI response. However, despite the marked difference in length of the source texts in the pilot corpus, the ratio of comments to source words in the two Faculties as shown in Table 3 was reasonably close (Fac1 = 1:130, Fac2 = 1:143), as was the ratio of categorized functions to source words (Fac1= 96, Fac2=106). For this reason, normalization was considered useful when comparing variables of grade band and Faculty.

Some source texts received inline (i.e., comments appended to source text) and summary (i.e., comments separated from the source text, often in response to overall criteria) WFI, and some received only inline WFI. All texts were moderated by a second marker. An overview of the pilot WFI corpus by module is given in Table 3.

[5] These included case-study reports, lab reports and empirical research reports.

Table 3

Pilot Dataset Overview

Faculty	Total source texts	High- band source texts	Middle-band source texts	Low- band source texts	Total source text words	Average number of in-text WFI comments	Average length of in-text WFI comments (words)	Average number of summary WFI comments	Average length of summary WFI comments (words)	Ratio of WFI comments to source words	Number of markers
1	3	1	1	1	25,502	18	13	3	129	1: 130	2
	3	1	1	1		35	13	10	83		2
2	3	1	1	1	11,768	22	11	0	0	1:143	2
	3	1	1	1		0	0	5	105		2

4.2 Sociodemographic Metadata

As part of the pilot corpus creation, we also collated, anonymized, and categorized student sociodemographic variables based on information reported by students on entrance to HE (such as date of birth, disability, nationality, ethnicity, home postcode), and also the course-related variables that describe their academic status (such as course, qualification on entry, level of study). These variables form part of the header metadata, along with WFI-related variables. Because of the small number of texts analyzed in this pilot corpus, there is insufficient data to look for correlations with many of the variables, including ethnicity. Also, we did not have access to information about the backgrounds of the academic staff who delivered the feedback. We focus on the Faculty in which the paper was written, and the grade obtained for the piece of work for which WFI is given, as seen in Table 3.

All papers were encoded in XML following TEI P5 standards (TEI, 2019), including standard header fields for written texts. Header metadata was automatically associated with each XML file from its CSV source using a custom-built script. Feedback categories in the main body were encoded by hand using a simple editor, and data extraction and counting processes were also done using a custom script[6].

It should be noted that source texts were marked anonymously, so none of the sociodemographic or course-related variables were directly available to markers. We cannot assume that markers were unaware of the identity of the student writer of the source text because opportunities to become familiar with content arise through, for example, giving formative feedback, WFI, or in-class discussions of an assignment. Equally, we cannot assume any influence from unconscious bias towards the student writer. We address this complex issue of potential mechanisms by which inequitable impact on certain groups of student writers based on sociodemographic factors occurs in the Discussion (Section 6.0 below).

4.3 WFI Framework Development

The WFI framework resulted from a multi-phase, cyclical construction process. In the first phase, two researchers reviewed existing frameworks (e.g., Hamp-Lyons & Chen, 1999; Hattie & Timperley 2007; Hughes, Smith, & Creese, 2015) and developed a working list of potentially relevant functions. In phase two, one researcher then amended the working list in light of the pilot data by: a) establishing a hierarchy of functions, and b) adjusting the nature of functions (or, elements), and, particularly, the branching sub-functions (or, attributes). At this point, elements were assigned attributes along three dimensions. For example, the element "critique" was assigned a *focus* attribute according to whether the focus was on "content" (1) or on "form" (2); each of these was then assigned an *orientation* attribute based on whether the critique was "specific" (A) or "general" (B). In the third dimension, the *aspect* of the critique was classified according to whether it related to the "present" (P) assignment

[6] For our planned larger dataset, to which multiple researchers of varying levels of experience will contribute, we will use a corpus tool to associate separately-held metadata with source texts. This will also allow categories to be preloaded to ease manual annotation, and it will enable automatic mark up of structural elements and various levels of tagging in the body text.

or “future” (F) work (as shown in Figure 1). This process yields a single code for each feedback function. For example, C1AP refers to critique (C) of content (1) that is specific (A) and present (P). This coding perhaps appears rather complex for manual annotation of a small corpus, but the branching nature will be invaluable when the framework is applied to a large corpus of several million feedback items.

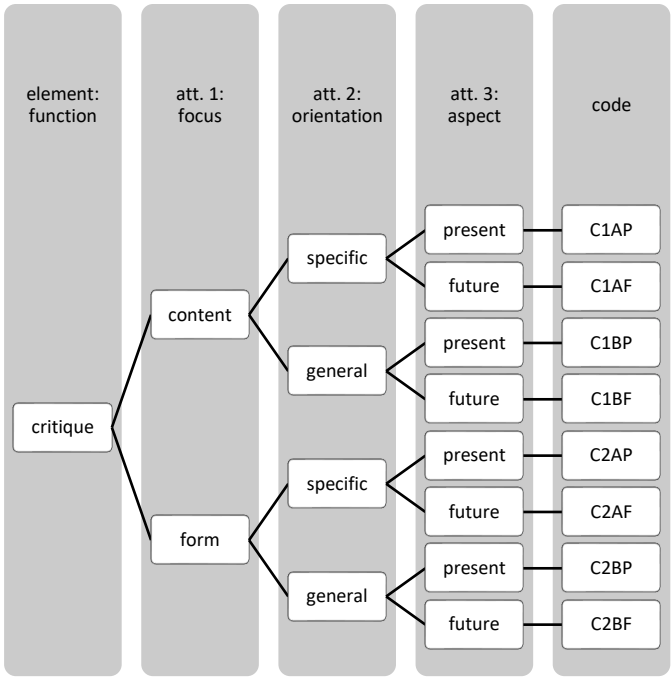

Figure 1. Example element and attributes tree for the function “critique.”

4.4 Feedback Function Annotation

In addition to the metadata in the header of each WFI file, in the body, we isolated distinct comments and annotated textual functions within these. Each comment is line-separated and structurally-tagged as either “summative” or “inline,” based on its position within the source text. Non-overlapping functions are annotated within each line-separated comment, for example:

> <comment type="summative"><P1A>An adequate attempt to use relevant articles and NICE guidelines to formulate a management plan</P1A> <C1B>however much of the work was superficial and the resulting management plan was not clear.</C1B></comment>

This summative comment has been encoded as P1A (specific direct praise) followed by C1B (general content critique).

> <comment type="inline"><Q1A>how will you manage the depression? How exactly will you, mange the LBP? How will you manage the obesity?</Q1A> <O2A>You have provided general ideas but it doesn't appear you have formulated a specific management plan for this patient</O2A></comment>

This inline comment has been encoded as Q1A (specific open query), followed by O2A (specific non-neutral observation).

All WFI comments were encoded in this way, which means that all words were assigned a single function based on the developed taxonomy.

4.5 Coding Processes

A first pass at annotation was then completed by one researcher using this hierarchy of functions. In the third phase, all annotated WFI was extracted into separate files based on code, which were discussed by the two researchers. Inter-coder reliability testing was not appropriate at this stage because significant changes to the hierarchy of functions and to the nature of the attributes (especially at the first level of *focus*) were considered necessary. Following these adjustments, both researchers completed a second pass at annotation independently, and results were compared based on complete or partial agreement over the boundaries of annotations. Although inter-coder reliability could not be objectively measured in this pass because coders were already familiar with the material, we wanted to gauge levels of agreement in terms of applying the revised taxonomy to the data. Agreement was calculated based on intersection, using the following formula (Alsop, 2015, p. 79):

$$agreement(A,B) = \frac{(\text{indices\_A } \text{Ç } \text{indices\_B})}{(\text{indices\_A } \text{È } \text{indices\_B})}$$

Three scenarios arose: full agreement, partial agreement, and disagreement, as exemplified in Table 4.

Table 4

Example of Three Types of Inter-Coder Reliability Intersection

	Agreement (boundary and function aligned)	**Partial agreement** (boundary fuzzy, some function intersection)	**Disagreement** (function mismatched)
R1	<comment type="inline" number=26><critique>**I still don't feel that you have provided a rationale for a management approach for this patient - I am unclear exactly what you are planning to do & why**</critique></comment>	<comment type="inline" number=26><critique>**I still don't feel that you have provided a rationale for a management approach for this patient - I am unclear exactly what you are planning to do & why**</critique></comment>	<comment type="inline" number=26><critique>**I still don't feel that you have provided a rationale for a management approach for this patient - I am unclear exactly what you are planning to do & why**</critique></comment>

	Agreement (boundary and function aligned)	**Partial agreement** (boundary fuzzy, some function intersection)	**Disagreement** (function mismatched)
R2	<comment type="inline" number=26><critique>**I still don't feel that you have provided a rationale for a management approach for this patient - I am unclear exactly what you are planning to do & why**</critique></comment>	<comment type="inline" number=26><critique>**I still don't feel that you have provided a rationale for a management approach for this patient -** </critique><observation>**I am unclear exactly what you are planning to do & why**</observation></comment>	<comment type="inline" number=26></observation>**I still don't feel that you have provided a rationale for a management approach for this patient - I am unclear exactly what you are planning to do & why**</observation></comment>
% overlap (29 tokens)	R1 = 29 R2 = 29 Intersection = 29 / 100%	R1 = 29 R2 = 17 Intersection = 17 / 59%	R1 = 29 R2 = 0 Intersection = 0 / 0%

Out of a total of 275 comments, full agreement was reached on 229, partial on 25 (with an intersection range of 32%–96%), and disagreement occurred in the case of 21 comments, giving a 0.83 probability of complete agreement (or 0.92 if fuzziness is allowed) at this stage. Where full agreement was not independently reached, comments were discussed until consensus was reached.

In some cases, function was difficult to extract from context. For example, the comment "yes however there is a lot the physio can do" could be interpreted in multiple ways, dependent on intonation and teaching context. For this reason, the category of "observation non-neutral" does not specify the direction of the implication (negative or positive), but instead indicates that it is loaded. In other cases, implication was not explicit, or even recoverable. Comments that appear to function purely to offer additional information are classified as "neutral" focus observation type.

Where functions are entwined within a comment that cannot be divided into standalone clauses, the primary function was encoded. For example, "this report lacks depth, critical analysis and coherence" is classified by the primary function critique-content (C1B): "depth" and "critical analysis," which can be interpreted at the level of ideas generally, are prioritized over "coherence," which could also refer to issues of general form (C2B). Out of 275 comments, only three examples of such entwining were identified.

5.0 Results

5.1 Feedback WFI Categories: Examples (RQ1)

RQ1: Can a framework of feedback functions be developed for our data that share a meaningful set of attributes?

The feedback framework includes five main categories (*praise*, *critique*, *advice*, *query*, and *observation*), as well as a catchall *unclassified* category. In terms of attributes, each umbrella category ("type" of function) can be described in terms of one of the two possible "focus" options and again as either *specific* or *general* "orientation." Examples are given in Table 5. The third attribute branch "aspect" (noted in Figure 1) is not included in this collapsed illustration because the majority of examples met the *present*, rather than the *future*, aspect criterion.

Table 5

WFI Function Taxonomy with Examples

Type	Focus	Orientation	Code	Example
praise	Ideas *expression of approval (content)*	specific *pinpoints an aspect of the assignment or writing process*	P1A	Links made between obesity, depression and LBP demonstrating an understanding of co-morbidities. good application here yes, very important
		general *refers to the overall assignment or writing process, or more widely*	P1B	The answer demonstrates a good understanding of theories, concepts and issues. You keep your answers to the point and framed around the context given to you in the scenario. It is clear you have read around this subject
	organization *expression of approval (surface form / adherence to genre conventions)*	specific	P2A	Your introduction clearly states the reasoning and the focus for this study. Abstract successfully summarises the project
		general	P2B	Relationships between statements and sections are easy to follow, and there is a sound, coherent structure. A very well constructed and written mini project. you write well and with clarity.

critique	content *error correction or recognition of a flaw (factual)*	specific	C1A	the plots for demonstrating various impacts (e.g. temperature values, 2D/3D simulations) on velocity profile and other flow characteristics are missing. This cant have been done by measuring SLR with a goniometer.
		general	C1B	Much of the presented work was descriptive and not explored in relation to the patient in the case scenario. Its all very confused
	form *error correction or recognition of a flaw (structural)*	specific	C2A	"wrap his head around" is an informal phrase not suited for academic work - better to say "understand" page number for direct quote used
		general	C2B	rather long-winded and could have been expressed more succinctly The text include some informality and indirect conclusion. Overall the essay suffers from poor structure and sentence structure.
advice	suggestion *recommendation for change, often highly hedged*	specific	A1A	Perhaps also point out that it is essential that HCPs look beyond the guidelines also to ensure you apply EBP You might also mention environmental conditions and different aircraft types that could be introduced could have shown combined effect here.

advice (con't)		general	A1B	You might find it useful to use a tool such as CASP when trying to critique the literature. I would recommend that you spend more time reading you work before submitting it to ensure it reads more smoothly.
	instruction *requirement for change, expressed using the imperative*	specific	A2A	you need to explain how you handled your data and what numbers you used in the statistical analysis. word this in relation to time/cost saving and user requirements explain this point
		general	A2B	You need to link everything together & build on your statements. you should be more precise in your analysis. you need to explain the points you make
query	open *a question with multiple possible answers, intended to encourage consideration*	specific	Q1A	How will you manage the obesity? Why? What do you mean by this?
		general	Q1B	[not applicable]
	closed *a question with limited possible answers, designed to elicit a factual response*	specific	Q2A	the whole nervous system or just the sciatic nerve? what constitutes a negative SLR? who read the goniometer & how was this recorded?

query (con't)		general	Q2B	[not applicable]
observation	neutral *a comment with no clear implications – often provides additional information or clarification*	specific	O1A	I think here you are highlighting some of the challenges facing HCP when trying to implement guidelines in patients with multiple pathologies. You are trying to explore how reliable these 2 measures are.
		general	O1B	Simulators can get out of step with the live environment if they are not updated to reflect modifications in the live environment, such as modifications to the tug, or new manoeuvring areas. This could result in negative training transfer There are a range of benefits from the training effectiveness perspective. Examples include: the ability to control the scenario, pause if required, plus collect data and replay the simulation for use in debriefs
	non-neutral *a comment loaded with non-explicit implications, usually recognition of an omission or a flaw*	specific	O2A	patient centred care is a little bit more than this You are starting to focus on the management of the patient here and not the benefits and challenges of the guidelines.
		general	O2B	some of this is sounding like a solution rather than analysis The project should be coherent & develop logically.
Unclassified			U	Thank you for submitting this assignment.

Although the main categories of this novel framework are informed by literature, unlike other frameworks, it is binary, symmetrical and exhaustive. The three levels of binary attributes (focus, orientation, and aspect) that branch from the main functional categories allow us to code our data in such a way that we can isolate variables at different levels of specificity and to cross-tabulate these with metadata. These attributes are symmetrical, which

allows accurate comparison across datasets. We were also able to classify all data using the framework, which included both in-text and summative types of feedback.

This framework describes the types of WFI that students receive. Effective feedback fills the gap between current and desired understanding/performance in a way that is specific to the relevant task or learning process and in a way that engages the receiver (Hattie & Timperley, 2007, p. 82). In our dataset, we assume that specific WFI is more likely to lead to future improvement in student work than general WFI because it is more likely to engage students at the level of ideas, so this distinction has been applied to all categories. We also assume that a balance of direct praise and critique alongside the potentially more open WFI types of advice, query, and observation will lead to improvement.

5.2 Classifying WFI Categories: Overall Patterns (RQ1)

The results in Table 6 show that the most common functions within comments are either *praise* (100 instances) or *critique* (105 instances) and that together they account for over 50% of all WFI, both in terms of number of instances (54%) and in terms of the length of those instances in words (51%).

Table 6

Breakdown of WFI Functions by Type, Focus and Orientation

funct.	**occurrence in comments**				**focus**	**occurrence in comments**				**orient.**	**occurrence in comments**			
	inst.	**% inst.**	**words**	**% words**		**inst.**	**% funct. (inst.)**	**words**	**% funct. (words)**		**inst.**	**% funct. (inst.)**	**words**	**% funct. (words)**
praise	100	26	887	23	ideas	91	91.0	791	89.2	spec.	68	68.0	499	56.3
										gen.	23	23.0	292	32.9
					orgs'n.	9	9.0	96	10.8	spec.	2	2.0	26	2.9
										gen.	7	7.0	70	7.9
critique	105	28	1083	28	content	67	63.8	747	51.2	spec.	51	48.6	484	44.7
										gen.	16	5.7	263	6.5
					form	38	36.2	336	31.0	spec.	6	5.7	70	6.5
										gen.	32	30.5	266	24.6
advice	51	13	574	15	suggest.	11	21.6	186	32.4	spec.	8	15.7	133	23.2
										gen.	3	5.9	53	9.2
					instruct.	40	78.4	388	67.6	spec.	27	52.9	254	44.3
										gen.	13	25.5	134	23.3
query	68	18	557	14	open	31	45.6	196	35.2	spec.	31	45.6	196	35.2
										gen.	0	0.0	0	0.0
					closed	37	54.4	361	64.8	spec.	37	54.4	361	64.8
										gen.	0	0.0	0	0.0
observation	54	14	831	21	neutral	17	31.5	291	35.0	spec.	14	25.9	202	24.3
										gen.	3	5.6	89	10.7
					non-neut.	37	68.5	540	65.0	spec.	30	55.6	451	54.3
										gen.	7	13.0	89	10.7
unclass.	1	0	6	0										
total	379		3938											

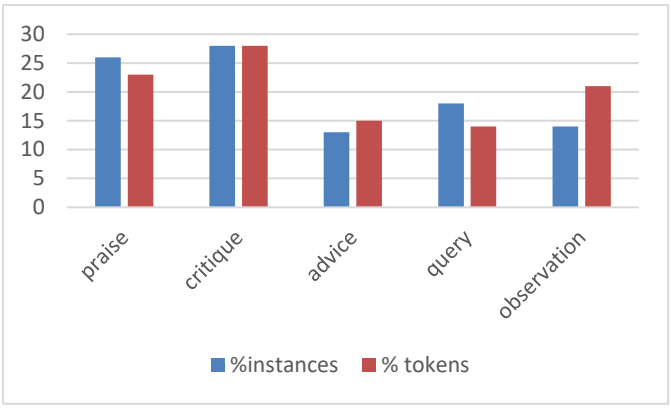

Figure 2. Breakdown of primary WFI function types by instances and words (%).

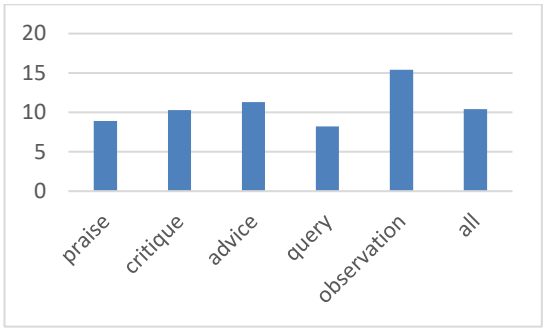

Figure 3. Average length per WFI instance by type (words).

The data in Table 6 can perhaps be more easily read in Figure 2 and Figure 3. Although *observation* types tend to include the longest comments (Figure 3), *critique* types are most frequent, followed closely by *praise* (Figure 2) in terms of both number of instances of each function and proportion in words of all WFI given. It is worth noting that *advice* is the smallest category of instances, and only *query* types are on average shorter in length. Moreover, the number of instances of *advice* is around half of each of *praise* and *critique*, suggesting that fewer than 15% of instances of WFI functions involve some kind of advice giving.

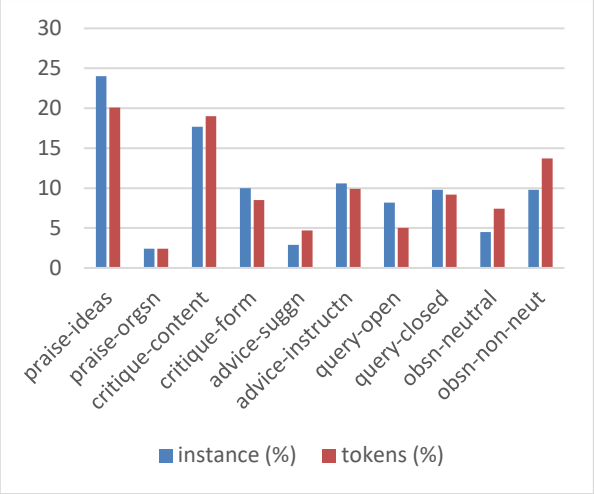

Figure 4. Breakdown of WFI type-focus by instances and words (%).

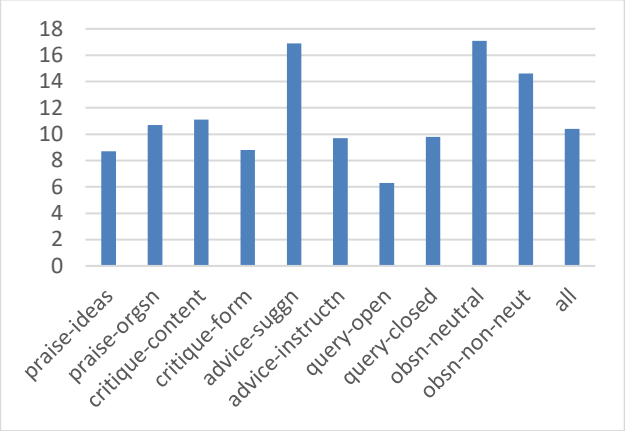

Figure 5. Average word length per WFI instance by type-focus.

Figure 4 shows that both *praise* and *critique* tend to focus more on substance (ideas and content) than structure (organization and form). In fact, instances of specific praise of ideas (P1A) account for 68% of all *praise* (56% words) and 18% of all instances of WFI function (13% words). Despite this high-frequency occurrence, the average length of each instance of P1A is shorter than the overall average (P1A=8.7; all=10.4, see Figure 5). Praise of organization is about one tenth as likely to occur as praise of ideas.

In a similar way, instances of specific critique of content (C1A) account for 49% of all *critique* (45% words) and 13% of all instances of feedback functions (13% words). The average word length of each instance of *critique* with this focus and orientation is slightly above average (C1A=11.1; all =10.4). This means that although academic staff tend to critique the substance of student work slightly less often than they praise it in terms of instances, they do so at slightly greater length.

The longest feedback comments are classified as *observation*, with an average of 15.4 words (Figure 3), which breaks down as 17.1 for neutral and 14.6 for non-neutral observations (Figure 5). This is followed by *advice* at 11.3 words (Figure 3) which breaks

down as 16.9 for suggestions and 9.7 for instructions (Figure 5). These two findings are important as they suggest that academic staff not only engage at length with the substance of student work but also provide directions for improvement.

Most of the advice given (78%) focuses on instruction (rather than suggestion), about two-thirds of which is specific in orientation. Overall, *advice* accounts for 11% of all instances of all WFI functions categorized and 10% of the overall word count (Figure 2), and all examples relate to the current assignment. This suggests that "telling" is a common mode of delivering WFI, and that academic staff tend to be explicit when doing so.

The "asking" type function, *query*, accounts for 18% of all function instances (and 14% of all words) (Figure 2). All of the queries are specific in orientation, and focus was divided between open and closed queries. Despite near parity in number of instances (46% open, 54% closed) (Figure 4), the average length in words per instance is markedly different: open = 6 words, closed = 10 words (35% and 65% total words, respectively) (Figure 5). It seems that academic staff put emphasis on querying information within the source text, and they invest more heavily in requesting concrete answers through query than in attempts to elicit general consideration.

As in the case of queries, the majority of observations were specific in orientation (82% instances, 78% words). Notably, over two thirds of all instances and words (68% and 69%, respectively) were non-neutral in focus (Table 6). This trend shows that much of the WFI delivered through observation is loaded in some way, which means that academic staff most commonly use observations to deliver other types of messages (such as critiques and suggestions), indirectly.

5.3 Classifying WFI categories: Patterns of Chaining (RQ1)

Most commonly, unique feedback functions align with unique clauses. Over three quarters of comments (212/275) were encoded with only a single function. In total, however, 379 instances of functions were encoded, which means that 167 instances of functions occurred within 63 comments. In these cases, either two or three functions operated as a "feedback sandwich" within a comment (Figure 6). For example:

> <P1A>Good point.</P1A> <O1A>Clinicians need to use clinical judgement alongside guideline recommendations.</O1A> <O2A>Even better if you had developed this further by arguing why you think this is important.</O2A>

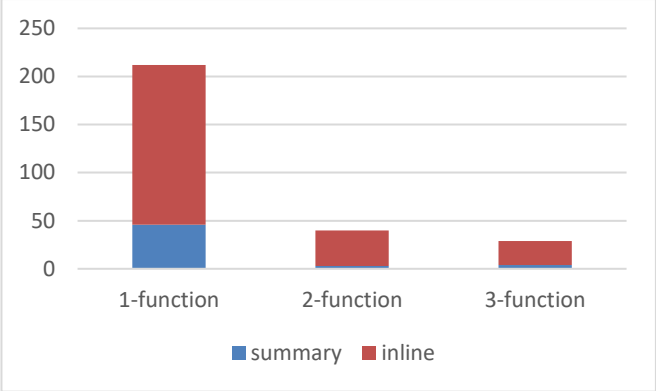

Figure 6. Function chains within comments (number of chains).

These multiple-function comments largely occurred inline (rather than in summative comments, as might have been expected): Out of 69 examples of two- and three-function chains, 62 occurred within inline comments (Table 7).

Table 7

Breakdown of Functions Within Comments

	1-funct.	2-funct.	3-funct.	total
summary	46	3	4	53
inline	166	37	25	228
total (number of chains)	212	40	29	281
total (instances of function)	212	80	87	379

No clear patterns of chaining were indicated in the three-function comment sandwiches, or if two-function comments were analyzed at the attribute level (see Figure 7). Some predictable patterns were, however, indicated at the element level in comments that contain two functions, such as a relatively higher frequency of some kind of *praise-criticism* (P-C) type sandwiches (see Figure 8), such as:

> <P1A>An adequate explanation of results, addressing the most of learning objectives, with analysis, synthesis and evaluation.</P1A> <C1A>However, the plots for demonstrating various impacts (e.g. mesh independence check, temperature values, unsteadiness) on velocity profile and other flow characteristics are missing.</C1A>

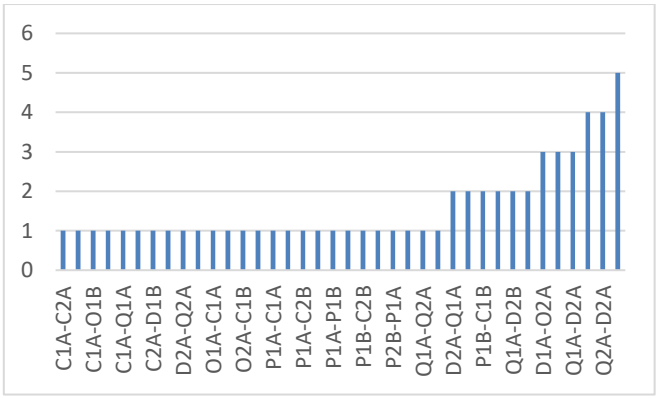

Figure 7. Occurrence of 2-function comments (attribute level).

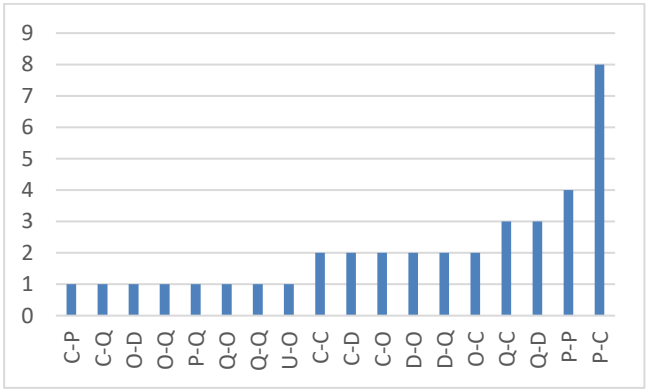

Figure 8. Occurrence of 2-function comments (element level).

These summary statistics from four modules in two Faculties give us a baseline against which other data can be compared. We now turn to comparisons across groups of students within this pilot data set in order to explore whether our classification enables us to identify crucial areas of difference.

5.4 WFI Categories (by Faculty) (RQ2)

RQ2: Can these functional categories be used to differentiate feedback to different groups of students?

The analysis of which faculties use which WFI functions shows a difference in usage. Overall, academic staff in Fac2 engaged much more often in praising and critiquing the source texts, whereas academic staff in Fac1 put more emphasis on query type WFI (Figure 9). Drilling down to the level of WFI orientation (Figure 11), however, reveals a slightly different picture. Although *praise* is more common to Fac2, academic staff in Fac1 more commonly give praise at the level of ideas specific to the current text.

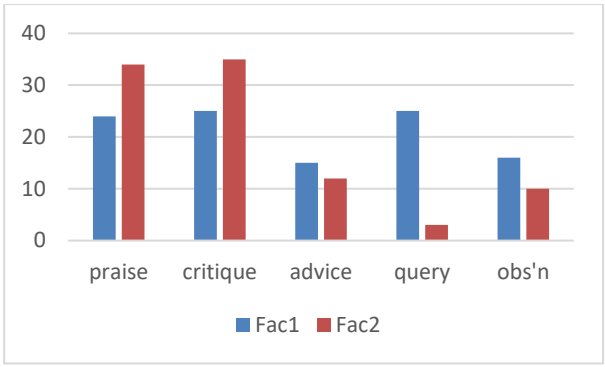

Figure 9. Distribution of functions by faculty (per 10,000 words) (type).

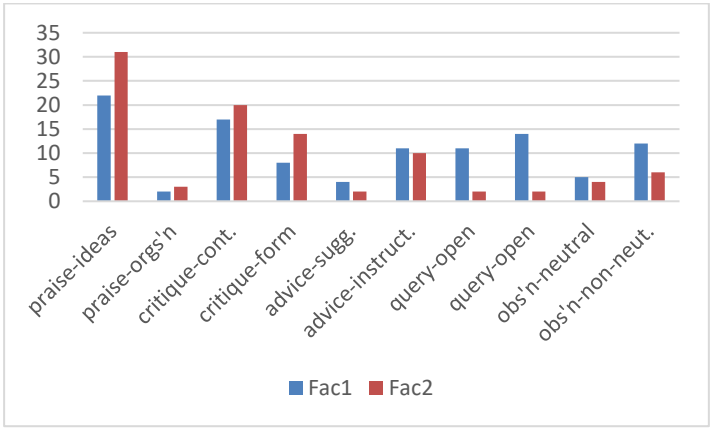

Figure 10. Distribution of functions by faculty (per 10,000 words) (type-focus level).

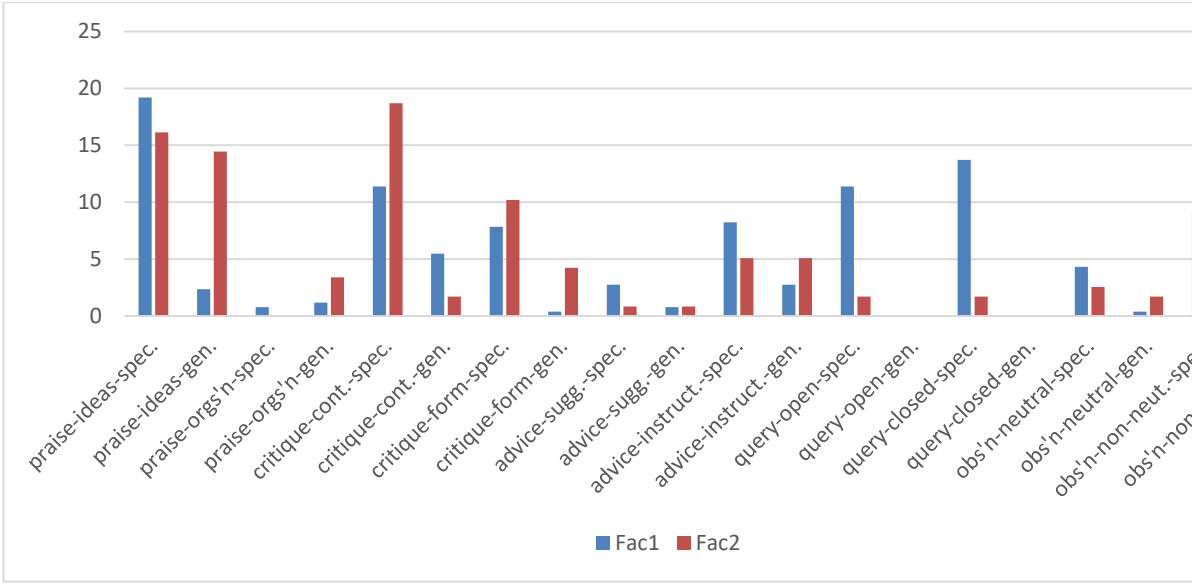

Figure 11. Distribution of functions by faculty (per 10,000 words) (type-focus-orientation).

5.5 WFI Categories (By Grade Band) (RQ2)

Overall, the lower the grade band of the source text, the more WFI comments received (low band=123, middle band=94, high band = 60). The number of WFI functions within these comments was likewise unevenly balanced (low band = 185, middle band = 119, high band = 74).

At the level of WFI function type, all texts received roughly equal amounts of *praise*, high-band texts received the least *critique*, and low-band texts received the most *advice*, *query*, and *observation*. This suggests that academic staff are attempting to engage in some form of dialogue, particularly with students who produce low-band responses. High-band texts received notably little observation or query of any type, which, somewhat surprisingly, may indicate a lack of dialogic engagement by staff. An interesting outcome of this analysis is the gap identified between the amount and type of WFI given to low-band texts compared to middle-band texts. For example, roughly the same amount of critique is delivered to both bands (Figure 12), and this tends to focus on content rather than form (Figure 13), but fewer queries with open focus are directed towards middle-band texts (Figure 13). Similarly, almost equal amounts of neutral focus observation were delivered to low- and middle-band texts, but more non-neutral observation was delivered to low-band texts (Figure 13). Examples of instructive advice within lower-band texts are also more frequently specific in nature than the more general instruction given to middle-band texts (Figure 14). Such differences underline the importance of looking at the type of WFI that is delivered to different groups of students across a full range of marks attained.

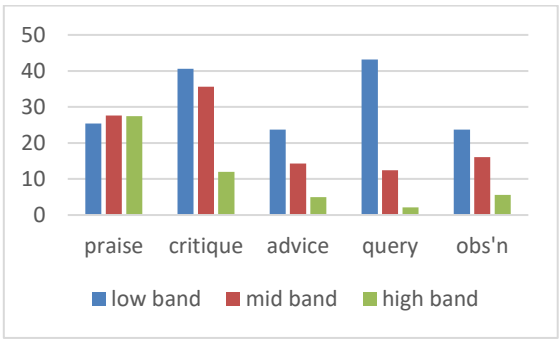

Figure 12. Distribution of functions by grade band (per 10,000 words) (type level).

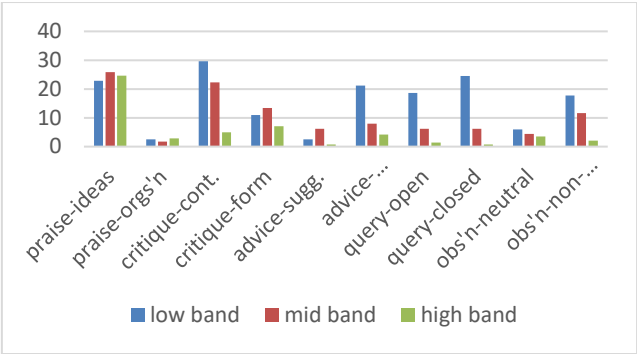

Figure 13. Distribution of functions by grade band (per 10,000 words) (type-focus level).

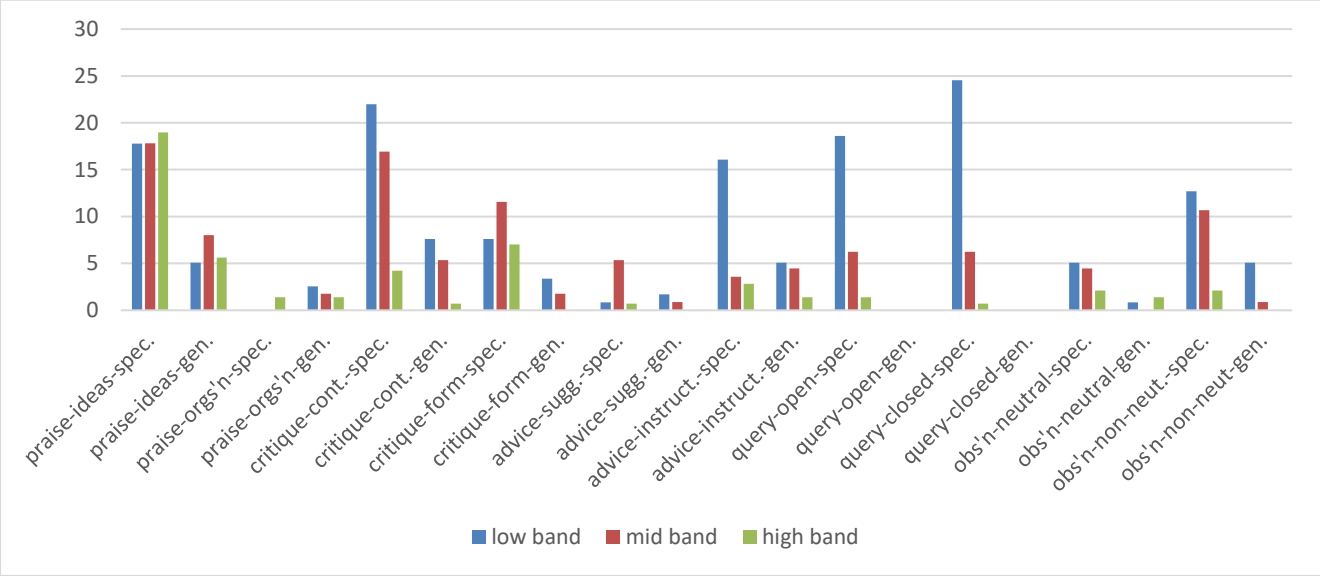

Figure 14. Distribution of functions by grade band (per 10,000 words) (type-focus-orientation level).

5.6 Summary of Results

The results show that all WFI comments can be classified using the framework developed. *Praise* and *critique* were found to be the most frequent type of comment, accounting together for 50% of the feedback given. Specific praise (P1A) accounted for 68% of all *praise*, while specific critique (C1A) accounted for 49% of all *critique*. Observations tended to be the longest comments (average 15.4 words). Analysis by Faculty shows two distinct styles, with Fac1 giving more *advice*, *query*, and *observation* than Fac2, where Fac2 gave more *praise* and *critique* than Fac1. Analysis by grade band indicates that there is greater dialogic engagement through *advice*, *query*, and *observation* in WFI to low-band texts compared to high-band texts.

6.0 Discussion

The small corpus developed and analyzed here incorporated two known contextual variables: discipline and grade. We know from student evaluations that students have given more favourable evaluations to feedback from Fac2, and our analysis shows that the model is able to point to differences in the types of feedback that may explain this. Similarly, the model is able to point to differences in the feedback given on texts of different grades. We know from the literature (e.g., Gardner, 2004) that more feedback tends to be given on lower-grade texts, and while our findings bear this out, they also suggest that the amount given on middle-grade texts is similar to that of low-grade texts.

Although the data was insufficient to focus on ethnicity, intersectionality, and attainment, these observations can be used to provide a baseline for further analyses. They have given us confidence that the model can be used to reveal how differences in types of WFI correlate with contextual variables. They also point to hypotheses about the positive ways WFI engages students in dialogue and the relative absence of WFI that facilitates closure of the loop (Carless & Boud, 2018). These themes will be explored further with a larger dataset where we will also be able to differentiate feedback to BAME and White students across disciplines and levels.

6.1 Engaging Students in Dialogue

We know that one of the ways to engage students is by enabling some form of dialogue through feedback (e.g., Hughes, Smith, & Creese, 2015), which is achieved via textual functions like *query*, *advice*, and *observation*. These dialogic devices form a core part of our functional taxonomy. We found that they are used differently and to different extents based on the variables we encoded. Generally, dialogic functions tend to be specific and instructive (either explicitly or through loaded implication), pointing students towards particular ways in which to improve a particular content-related aspect of a text. Some types of WFI commonly prompt the writer through direct address (such as the *advice* "You could have put the search into your appendices to expand on this"), and others tend to do so more indirectly using third person (such as the *observation* "This would have been clearer in a table"). Much less common are instances when the marker demonstrates engagement with the ideas within the paper through offering additional information (non-neutral observations), or asking non-closed questions (open queries). In these cases, it is not clear whether the comments affect the grade or the quality of the submitted paper, or whether they are intended as *praise* or *critique*, or simply as a means of extending the student writer's work/thinking. They do, however, offer a way of reaching out to the student writer. Capturing and distinguishing different levels and modes of conversation within WFI allows such attempts to foster engagement to be gauged.

The process of understanding what a marker is doing through WFI (when a comment is taken as the unit of analysis) is complicated by the occurrence of multiple functions within a comment. We noticed, for example, that the overall instructive nature of a comment was amplified when dialogic functions co-occurred in "feedback sandwich" chains. The act of double- or even triple-teaming functions within comments was frequently used to clarify

expectations, and commonly occurred in *query*/*advice* (or *advice*/*query*) chains, such as: "so apply this to your patient. why is this relevant?", "provide ref to support this statement. how do you know this?", "why? Explain." In these examples, the instructive *advice* makes explicit the call to action implied by the *query* element. It also, however, potentially limits dialogue (more so than if the presence of query was taken as a standalone function), as through the chaining of functions the student is directed rather than prompted. In this way, both single and multi-chained dialogic functions can mirror *praise* and *critique* functions, which also tend to focus on specific aspects of the current text. Such methods of delivering WFI make explicit appropriate action, which may be more or less important to a student dependent upon previous educational/cultural experience, but they also potentially limit recipient agency and engagement. Looking at chains of functions offers an additional nuance to understanding engagement.

Some differences in extent of engagement through WFI began to emerge when findings were organized by variables of Faculty and grade band. Overall Fac1 employed dialogic functions to a greater extent than Fac2. Fac2 instead delivered specific WFI about content more often through non-dialogic text, particularly *critique*. Across the dataset, most of the asking/telling dialogue was directed towards low-band texts; conversation decreased as mark increased, notably so around middle-band texts. The only function most relatively frequent in WFI delivered to high-band texts was praise of ideas in the current text, indicating that the highest attaining texts receive by far the lowest number of engagement-prompting WFI comments in this dataset.

Importantly, we noticed differences in distribution depending on the level of detail to which the taxonomy was applied. An overall higher frequency of *praise* within one Faculty, for example, obscured the higher frequency of particular types of *praise* within another. Identifying when and how such functions occur, in response to texts from which groups of students, allows us to measure the extent to which academic staff are (or are not) asking questions or making suggestions that prompt open consideration, or that could be applied beyond the current text. Moreover, the need for a fine-grained taxonomy (allowing for complex attributes) was highlighted. The ability to extract such patterns at multiple levels is important because we expect observation, query, and advice functions to play an important role in promoting student engagement, which should lead to greater attainment.

6.2 Enabling Closure of the Loop

Perhaps the most important driver of engagement is enabling students to close the feedback loop (Boud & Malloy, 2013; Carless & Boud, 2018); this process of actioning the WFI received has a direct relationship to attainment. Our taxonomy allows for present[P]/future[F] aspect to be attributed to identification of WFI type, focus, and orientation, enabling us to describe whether the WFI relates to the current assignment or to future work. One recognized method of enabling students to action comments is the use of ipsative feedback in relation to progress (e.g., Hughes, Smith, & Creese, 2015, p. 1081). In our dataset, we identified only two instances that could be described as ipsative feedback, both of which occurred within praise of the current assignment: "More relevant argument is presented in this paragraph" and

"more good points made in this section." Although *praise* was one of the most common functions, little of it was benchmarked, and none was applied to future work.

The future[F] aspect attribute identifies WFI that explicitly facilitates closure of the feedback loop. In our dataset, only one such example was identified, and it occurred within a three-function chain: praise-ideas-general-present (P1BP)-> critique-content-general-present (C1BP)-> advice-instruction-general-future (A2BF):

> P1BP>This was a good answer.</P1BP> <C1BP>However it suffered from too much padding; covering information that wasn't relevant to the question.</C1BP> <A2BF>In future try to be more direct in your answer to the scenario.</A2BF>

Some examples identified as of general orientation may provide transferable advice, such as the observation: "You are always better to explain the descriptive stats first & then use the inferential stats to decide if the findings are significant". Likewise, some general orientation comments may be transferable to the process of composing other assignments, such as the instructive advice "Ensure you understand what your RQ is and what you are trying to explore when you design your study", or the non-neutral observation "Points made should be explained, linked & supported by the evidence." These examples do not, however, explicitly point students towards how they can apply the WFI received to other work. By distinguishing the present/future aspect attribute of WFI, we can gain a more complex understanding of the role of WFI in the context of an ongoing academic journey and how this may impact differently on different groups of students.

7.0 Conclusion

Our first contribution to better understanding the relationship between WFI, the sociodemographic and academic characteristics of the recipient, and the grade attained by the recipient was to establish a means of categorizing function/s at the level of comment. Findings from this pilot indicate that WFI comments tend to function in a non-dialogic and non-ipsative way (such as *praise*, *critique*, and *advice* that is instructive related to the current assignment), which is in line with a recognized trend for feedback over feedforward (e.g., Hughes, Smith, & Creese, 2015, p. 1092). Our second contribution was to establish that the categories can be used to differentiate feedback to different groups. In our dataset, differences in usage of WFI were established in relation to the variables of Faculty and grade attained. This trend may have implications for facilitating closure of the "feedback loop," which in turn may impact student engagement and sense of belonging. This factor provides one avenue for investigating, and ultimately addressing, the acknowledged attainment gap between different groups of students, particularly groups with different self-identified ethnicities.

8.0 Directions for Future Research

The next stage in this project will be to apply the principles of categorizing feedback and using these categories to differentiate feedback to different student groups to specifically investigate the BAME/White attainment gap in UK HE. The next stage in analysis will be to identify the linguistic characteristics of the functions identified in order to automate the annotation process across the complete dataset (c. 80 million words), which will include WFI

delivered to assignments from all faculties, across all levels of undergraduate study in two UK institutions. We intend to use machine learning of the annotations to do so, based on a "gold standard" sub corpus created following full inter-coder reliability protocols. We also intend, in parallel, to perform automatic linguistic analyses on this larger dataset to triangulate the validity of the initially qualitatively identified functions against quantitatively identified patterns. Our approach to analysis will be exploratory, and we anticipate that features such as pronoun use, positive and negative appraisal features and tense aspect markers will be salient. The results of both taxonomies will be mapped onto the sociodemographic (including self-identified ethnicity) and academic variables (including mark attained) associated with the student recipient of each feedback comment, and results will be cross-tabulated. Looking at the data in this way will allow us to fully test the hypothesis that a relationship between WFI language, function, mark attained, and recipient ethnicity exists. It remains to be seen whether this relationship is circular in nature. We believe, however, that interrogating authentic data in the way proposed will enable researchers and also policy makers to better understand the complex relationship between ethnicity, attainment, belonging, and retention.

Author Biographies

Dr. Siân Alsop is Research Fellow in Global Learning at Coventry University, UK. Her work explores equity and attainment in education using corpus linguistic and data visualisation methods. She has worked on large corpora of written and spoken academic discourse, including the British Academic Written English (BAWE) corpus and the Engineering Lecture Corpus (ELC).

Dr. Sheena Gardner is Professor of Applied Linguistics at Coventry University, UK. Her research uses functional, corpus, and genre-based approaches to investigate the academic English of educational assessment. She worked with Siân Alsop on the BAWE corpus project. They also co-edited *Systemic Functional Linguistics in the Digital Age* (Equinox, 2016).

References

Alsop, S. (2015). The pragmatic annotation of academic lectures: An exploratory study applied to Engineering disciplines (unpublished doctoral thesis). Coventry University, Coventry, UK.

Boud, D., & Molloy, E. (Eds.) (2013). *Feedback in higher and professional education*. London and New York: Routledge.

Cabinet Office (2017). *Race disparity audit: Summary findings from the Ethnicity Facts and Figures website*. Retrieved from https://www.gov.uk/government/publications/race-disparity-audit

Carless, D., & Boud, D. (2018). The development of student feedback literacy: Enabling uptake of feedback. *Assessment and Evaluation in Higher Education*, *43*(8), 1315–1325. doi: https://doi.org/10.1080/02602938.2018.1463354

Connor, H., Tyers, C., Modood, T., & Hillage, J. (2004). *Why the difference? – A closer look at higher education, minority ethnic students and graduates*. DfES Research Report RR552. Retrieved from https://www.employment-studies.co.uk/resource/why-difference

Doku, A. (2018) Let's fix the Black attainment gap [Blog post]. Retrieved from https://wonkhe.com/blogs/fix-black-attainment-gap/

Gardner, S. (2004). Knock-on effects of mode change in academic discourse. *Journal of English for Academic Purposes*, *3*(1), 22–38.

Hamp-Lyons, L., & Chen, J. (1999). An investigation into the effectiveness of teacher feedback on student writing. *English Language Teaching and Learning 3*, 207–219.

Hattie, J., & Timperley, H. (2007). The power of feedback. *Review of Educational Research*, *77*(1), 81–112.

HESA (2018). Fields required from institutions in all fields: Ethnicity. Retrieved from https://www.hesa.ac.uk/collection/c18051/a/ethnic

Hughes, G., Smith, H., & Creese, B. (2015). Developing a feedback analysis tool to explore feed forward in modularised programmes. *Assessment & Evaluation in Higher Education*, *40*(8), 1079–1094.

Kluger, A. N., & DeNisi, A. (1996). The effects of feedback interventions on performance. *Psychological Bulletin*, *119*(2), 254–284.

Lee, K. Y. (2013). A genre analysis of written academic feedback (unpublished PhD thesis). University of Birmingham, Birmingham, UK. Retrieved from http://etheses.bham.ac.uk/4722/

Mavelli, L. (2014). Widening participation, the instrumentalization of knowledge and the reproduction of inequality. *Teaching in Higher Education*, *19*(8), 860–869.

McCormick (2015). The real effects of unconscious bias in the workplace. Retrieved from http://www.kenan-flagler.unc.edu/~/media/Files/documents/executive-development/unc-white-paper-the-real-effects-of-unconscious-bias-in-the-workplace-Final

Office for Students (OfS) (2018a). Data and analysis: Differences in student outcomes. Retrieved from https://www.officeforstudents.org.uk/data-and-analysis/differences-in-student-outcomes/

Office for Students (OfS) (2018b). National Student Survey. Retrieved from https://www.officeforstudents.org.uk/advice-and-guidance/student-information-and-data/national-student-survey-nss/

Pitt, E., & Norton, L. (2017). 'Now that's the feedback I want!' Students' reactions to feedback on graded work and what they do with it. *Assessment & Evaluation in Higher Education 42*(4), 499-516

Poe, M., Nastal, J., & Elliot, N. (2019). Reflection. An admitted student is a qualified student: A roadmap for writing placement in the two-year college. *The Journal of Writing Assessment*, *12*(1). Retrieved from http://journalofwritingassessment.org/article.php?article=140

Richardson, J.T.E. (2015). The under-attainment of ethnic minority students in UK higher education: What we know and what we don't know. *Journal of Further and Higher Education*, *39*(2), 278–291.

Tai, J., Ajjawi, R., Boud, D., Dawson, P, & Panadero, E. (2018). Developing evaluative judgment: Enabling students to make decisions about the quality of work. *Higher Education 76*(3), 467-481. https://doi.org/10.1007/s10734-017-0220-3

TEI (2019). Releases of the TEI Guidelines. Retrieved from https://www.tei-c.org/release/doc/tei-p5-doc/en/html/TitlePageVerso.html

University of Edinburgh (2017). Corpus analysis of student feedback. Retrieved from http://www.ed.ac.uk/institute-academic-development/learning-teaching/funding/funding/previous-projects/year/march-2017/student-feedback-analysis

UUK & NUS (2019). Black, Asian and minority ethnic student attainment at UK Universities: #closingthegap. Retrieved from https://www.universitiesuk.ac.uk/policy-and-analysis/reports/Pages/bame-student-attainment-uk-universities-closing-the-gap.aspx