Does EAP affect written L2 academic stance? A longitudinal learner corpus study

Peter Crosthwaite [a, *], Kevin Jiang [b]

[a] School of Languages and Cultures, University of Queensland, Australia
[b] School of Foreign Language Education, Jilin University, China

ARTICLE INFO

ABSTRACT

This study explores the longitudinal development of L2 academic 'stance' features resulting from instruction in English for Academic Purposes (EAP) at a university in Hong Kong. We analysed the frequency and wordings of *hedges, boosters, attitude markers* and *self-mention* within a 205,682 word longitudinal corpus of essays and reports collected over a semester's instruction via pre-, mid- and post-instruction submissions, alongside data on submission grade. Data was analysed for frequency and wording differences alongside mixed-effect models to confirm the impact of instruction on the data. Results show significant longitudinal variation in the frequency of hedging, boosting, marking attitude and self-mention devices as the result of instruction, with a rise in the use of hedging and an overall reduction in the use of boosting and self-mention, serving to leave students with a more careful, narrower, less polarising and less personal range of expressions with which to convey their stance over time. We also present longitudinal genre-specific effects on stance features between essays and reports, and show how a longitudinal increase in hedges and boosters results in texts that receive a higher grade from teacher-raters. Our findings recommend explicit instruction of stance features as crucial in raising students' awareness of how to achieve persuasive academic writing.

© 2017 Elsevier Ltd. All rights reserved.

1. Introduction

When transitioning to tertiary education, a crucial issue students in foreign language (L2) contexts face is the need to adopt an academic register, which, for many such students, is a considerable challenge (Hyland, 2016). In the context of the present study (Hong Kong), students lack the L2 proficiency to fully participate in tertiary education, often failing to understand lectures or write required coursework assignments (Evans & Morrison, 2011; Bruce & Hamp-Lyons, 2015). For L2 students (and to some extent, even students where the tertiary medium of instruction [MOI] is their native language), academic discourse is often regarded as "an alien form of literacy [… with] many students arriving at university thinking they have landed on Mars" (Hyland, 2016, p. 246). Many universities in both English and mother-tongue MOI contexts therefore provide language enhancement courses in English for Academic Purposes (EAP) for freshmen L2 students, aiming to develop students' essential knowledge base of the general rhetorical features and structures of academic discourse. This is intended to allow students to enculturate into academic life as they progress through their studies.

* Corresponding author.
  E-mail addresses: drprc80@gmail.com (P. Crosthwaite), kevinjiangfeng@gmail.com (K. Jiang).

A key aspect of 'successful' academic discourse is that of the writer's ability to construct, support, defend and justify an argument on a given topic (Hyland, 2000, 2005; Lee & Deakin, 2016; Wingate, 2012). A now longstanding EAP research area is that of the determination and analysis of the linguistic features involved in successful argumentative discourse, focusing on the particular linguistic *stance* devices, or "writer-oriented features of interaction" (p. 178) used by writers to continually engage the reader (Biber, 2006; Hyland, 2005). Studies of the stance features involved in academic writing are now numerous (e.g. Hyland, 2000; Jiang, 2015; Jiang & Hyland, 2016) and have led to the creation of new EAP course materials focusing specifically on these features (e.g. Chang, 2010), as well as the bottom-up, linguistic derivation of 'successful' academic texts (Lee & Deakin, 2016). For pedagogy, Chandrasegaran (2013) has shown that deconstruction of sample texts by stance features alongside explicit teaching of rhetorical grammar for realising desired genre goals have improved students' stance-making abilities, while Tribble and Wingate (2013) have suggested that corpora should be exploited for student development of linguistic knowledge and skills needed for academic argumentation.

Yet, while a considerable amount of studies have compared stance-taking practice between L1 and L2 writers (e.g. Hinkel, 2005; Lee & Deakin, 2016), less is known regarding how L2 learners develop stance expression as the result of EAP instruction. Addressing this gap would open a window into the overall effectiveness of EAP instruction in producing successful academic writing. Given the scale of student numbers, text types and linguistic information required for an incisive and detailed appraisal of EAP and stance making, this paper reports on a longitudinal corpus-based study of stance presentation during in-sessional L2 EAP written production. We aim to determine how EAP instruction over time affects L2 development of stance features in EAP written essays and reports. We also provide accompanying data on learner variables so as to delineate the impact of EAP instruction from potential learner-internal factors influencing L2 stance development, as well as identify the stance features involved in teacher-raters' positive appraisal of 'successful' L2 academic texts. We begin by outlining the rationale behind longitudinal learner corpora for the investigation of learner stance development, before providing an overview of the linguistic features under investigation.

2. The need for longitudinal EAP corpus analysis

Corpora, or "a principled collection of language materials, spoken or written, compiled into an electronic database for the purpose of linguistic analysis" (Park, 2014, p. 27, see also; Biber, Conrad, & Reppen, 1998; Sinclair, 2004) are now considered to have multiple affordances (Leńko-Szymańska & Boulton, 2015) for language learning and teaching, with the EAP field no exception. As mentioned, corpus-based explorations of stance features are now numerous (e.g. Hyland & Tse, 2004; Hyland, 2005). However, there remains a necessity to compile *longitudinal* corpora, so that "a thorough, in depth examination of development over time can be made" (Park, 2014, p. 39). The end goal of this endeavour is to generate data that allows teachers to "define areas that need special attention in specific contexts and at different levels of competence, and so devise syllabi and materials" (Gabrielatos, 2005, p. 6). Such corpora therefore provide insights into the language learning *process* rather than the end product, supported by fine-grained analysis of authentic L2 production.

Longitudinal corpus-based studies into EAP remain relatively scarce compared to their pseudo-longitudinal (i.e. by L2 proficiency cross-section) counterparts such as the Cambridge Learner Corpus (Nicholls, 2003) or the International Corpus Network of Asian Learners of English (ICNALE; Ishikawa, 2013). This is primarily because collecting longitudinal data "is a real challenge, as it is both time consuming and requires much planning ahead [… with] few research terms collecting such data types" (Meunier, 2016, p. 381). However, Granger, Gilquin, and Meunier (2016) note that longitudinal corpora are "showing a slow and steady rise" (p. 2) with projects such as the Longitudinal Database of Learner English (LONGDALE, Meunier & Littré, 2013), recent edited volumes including projects using longitudinal corpora (Castello, Ackerly, & Cocceta, 2016), as well as a special issue of the *Modern Language Journal* (Hasko & Meunier, 2013) covering such studies.

However, despite the recent popularity of longitudinal corpus methodology, data from EAP learners in the Asian context is relatively underexplored, as researchers have been slower to adopt the construction of such corpora, and existing longitudinal corpora lack much in the way of Asian L2 EAP discourse. For example, in what is currently the largest longitudinal learner corpus worldwide - the 32 million-word EFCamDAT longitudinal corpus (Geertzen, Alexopoulou, & Korhonen, 2013), of which nearly 19% of labelled as 'Chinese' - each submission has an average length of just seven sentences, much shorter and less specialised than would be seen in the majority of EAP production. A recent review of other longitudinal corpora (Meunier, 2016) has pointed out 11 such corpora, of which 10 contain data only from European sources. One exception is found in Crosthwaite (2016), whose study on Asian L2 EAP writing using Biber's (1988) multidimensional analysis determined that EAP instruction resulted in longitudinal linguistic variation in the direction of the established norms of an academic register. This variation included increased an emphasis on nominalisation and more careful, hedged, presentation of stance, and targeted written corrective feedback was highlighted as one of the key pedagogical advantages of this type of analysis.

However, the multidimensional approach in that and other studies takes a snapshot of the entirety of the linguistic features of student production and compares it against established statistical norms for a given register — in this case, academic discourse. This approach may be akin to cracking a nut with a sledgehammer, in that the nuance or power of the writer's presentation of the epistemic and rhetorical values as encompassed in the specific wordings and function used by the writer may be lost without a more fine-grained analysis of the actual linguistic features used by the writer. The key with

longitudinal data is to determine whether EAP instruction, over time, has an effect on the writers' usage of such features in line with their increasing awareness of the conventions of academic discourse, and any findings would help to make visible the hidden curriculum of linguistic features that constitute raters' positive appraisal of performance, but are not explicitly taught on current L2 English curricula (Legg, 2016).

We now outline how a writer's selection of stance features is constitutive of their ability to construct successful − and academic − argumentative discourse.

3. Stance in academic writing

Rather than dry and impersonal, academic writing is a persuasive endeavour, saturated with the perspectives of the writer on the source material, while anticipating readers' imagined reactions to those views (Aull & Lancaster, 2014; Hyland, 2002, 2005; Lee & Deakin, 2016). Stance is something of a catch-all yet elusive concept, referring to the ways that writers project themselves into their texts to offer authorial views toward the content of their talk. Studies on undergraduate student writing show that the ability to express a relevant and plausible stance is a crucial indicator of writing quality and development (Aull & Lancaster, 2014; Lancaster, 2016; Lee & Deakin, 2016; Qin, 2014; Zhao, 2013). Zhao (2013), for instance, quantitatively shows that the appropriateness of authorial voice a student writer projects into a text has a positive relation with the rating of its quality. Moreover, Nesi and Gardner (2012), Miller, Mitchell, and Pessoa (2014) and Llinares and Dalton-Puffer (2015) have further demonstrated that students also find it an additional challenge to express an appropriate and relevant stance in line with the discursive conventions of different genres and task types they encounter in disciplinary courses.

Other concepts have also been used to describe stance-taking practice and means, such as *evaluation* (Hunston & Thompson, 2000) and *appraisal* (Martin & White, 2005). Hunston and Thompson (2000) talk of "the values ascribed to the entities and propositions" evaluated (p. 5) while Martin and White (2005) relate appraisal additionally to "those means by which writers more indirectly activate evaluative stances and position readers to supply their own assessments" (p. 2). However, Biber (2006) and Hyland (2005) take a more pedagogically-oriented perspective to these interactional resources, informing the way students can be sensitized to this rhetorical ability. In addition, their analytical approaches allow for a more automated corpus annotation process than the thorough manual reading required under *evaluation* and *appraisal*, and so we specifically draw on their perspective for this EAP study. For Biber (2006), stance expresses a writer's "personal feelings and assessments", including "attitudes that a speaker has about information, how certain they are about its veracity, how they obtained access to the information, and what perspective they are taking" (p. 87). More centrally, Hyland (2005) suggests stance includes three main components: *evidentiality*, *affect* and *presence* (p. 178). These relate to three important rhetorical questions that academic writers may bring to texts about any statement:

> "How certain do I want to be about this?
>
> What is my attitude towards it?
>
> Do I want to make myself prominent here?" (Hyland, 2016, p. 248).

In line with Biber (2006) and Hyland (2005), we see stance as the writer's expression of epistemic assessment, personal attitudes and self-presence. Notably, stance under this framework is achieved by the appropriate selection of *hedges* and *boosters*, *attitude markers* and *self-mentions*.

Hedges are those linguistic devices which show a writer's uncertainty and the decision to withhold complete commitment to a proposition, "allowing information to be presented as an opinion rather than accredited fact" (Hyland, 2005, p. 178), particularly when the writer's existing knowledge is vague (Hinkel, 2005). In contrast, boosters work to increase epistemic commitment to the writers' propositions (Hyland, 2005; Lee & Deakin, 2016), thus "functioning to narrow the discursive space and authoritatively steer the reader toward the writer's views" (Lancaster, 2016, p. 19). Although both features take on various forms, such as lexical words, phrases and clauses, and project important interactional meanings in academic writing, research has shown that their appropriateness in specific contexts crucially depends on communicative topic, purpose and the norms of a particular discourse community (Aull & Lancaster, 2014; Biber, 2006; Hinkel, 2005; Hyland, 2002; Nesi & Gardner, 2012). For example, Aull and Lancaster (2014) found in their study that advanced academic writers privilege caution and hedged claims over certainty in the discussion of research subjects, while first-year undergraduates tend to use more boosters and fewer hedges that achieve the opposite kind of stance in their writing of general argumentative texts.

Although attitudinal markers and self-mentions project writers' explicit presence and effectuate affective persuasion (Hyland, 2002), they are not always constitutive of successful academic discourse. For example, Lee and Deakin (2016) found in their study of A- and B-level L2 undergraduate essays that these two forms were the least frequently used interactional resources, and that A-level essays made less use of them than B-level texts. Lee and Deakin believe that these two most overt interactional forms signify subjectivity rather than objectivity, and thus academic writers typically refrain from explicitly marking personal attitudes and presence in texts. Similarly, in Lancaster's (2016) study, high-grade papers for both the social sciences and humanities feature significantly less use of self-mentions and a greater use of boosters than those seen for other

disciplines, noting that the method of expression of stance appears to have a critical impact on the success of academic writing.

The above section outlines how the aspects of stance that student writers project play a role in shaping the judgments of readers and educators in terms of writing quality. What is still uncertain is the extent to which EAP actually affects students' ability to construct and develop a register-appropriate and engaging stance, as evidenced by the longitudinal incorporation of said stance features into their writing as a result of instruction. Quantitative analyses of the longitudinal impact of EAP instruction on the production of stance features are therefore crucial in determining the overall usefulness of EAP as an enterprise for course administrators, teachers and students alike.

4. The study

In this study we investigate the longitudinal development of stance features (namely hedges, boosters, attitude markers and self-mention) from the beginning to the end of a semester's tertiary EAP provision in a longitudinal corpus (the HKU-CAES corpus). It addressed the following research questions:

RQ1) *Is there any variation in the longitudinal distribution and wording of stance features in L2 student written production pre-, mid- and post-EAP training?*
RQ2) *Is there any evidence to suggest EAP instruction results in genre-specific variation in stance features between L2 student essays and reports pre- to post-instruction?*
RQ3) *Which stance features are considered as constitutive of 'successful' academic writing as evidenced by the grade assigned to a given text by teacher-raters, and is there any longitudinal impact on the use of such features?*

The HKU-CAES learner corpus is originally described in Crosthwaite (2016), and is summarised for the reader here with additional information about the specifics of the EAP instruction related to stance.

4.1. Participants

The data is collected from freshman undergraduates sitting their initial EAP course ('*Core University English*') at a Hong Kong university, with the course lasting 14 weeks in duration (9 weeks of instruction, then 4 weeks for feedback, revision and examinations). Data is collected longitudinally at three data points, pre-course (Week 1, data point 1), immediately after the instructed part of the course (Week 10, data point 2) and the final written assessment (Week 14, data point 3). Data at a minimum of three points allows researchers to fit a developmental line for the derivation of linear progression, U or reverse-U shaped behaviour (Meunier, 2016). All data was collected as part of the students' regular EAP curriculum. In total, 87 participants across five EAP class groups provided data, with the same participants submitting data at each data point. Although attrition is a serious concern with longitudinal corpora (Meunier, 2016), only two students did not submit data at the second and third data point. 3 of the 5 class groups were taught by one teacher, with the remaining 2 taught by another (neither class was taught by the researchers). Both teachers had over ten years' experience teaching EAP, with one an L1 Cantonese female and the other an L1 English-speaking male. A questionnaire was administered prior to data collection for recording demographic information. (Table 1). All student participants were between 18 and 19 years of age, an equal mix of men and women. The students' L2 proficiency ranges from IELTS 6.5–8, with 6.5 being the minimum requirement to enter the university. The average band score was 7 across the cohort, with no one class significantly higher or lower in L2 proficiency than any of the others.

Table 1 shows the vast majority of participants are monolingual L1 Cantonese (with a smaller number of L1 Mandarin speakers), who have never previously lived abroad, hold an upper-intermediate level of L2 English, and with a 70%–30% split between hard and soft majors. The study was given ethical clearance by the participating institute and all students gave consent for their data and grade information to be used.

4.2. Data sample and collection conditions

The corpus sample size at each data point (henceforth DP) is summarised in Table 2. In terms of the conditions for data collection, DP 1 constituted a hand-written 'diagnostic writing task' (800 words), which was not assessed but was used as part of an in-class activity where teachers provided feedback on L2 errors made. Students were asked in the next session (week 2) to attempt to correct some of the errors highlighted as well as correct some of the errors of the person sitting next to them. The data collected at DP 2 (800 words) was word-processed and done out of class. If a student took the 'essay' question at DP 1, they obligatorily took the 'report' question at DP 2, and vice-versa. Teachers provided more detailed written corrective feedback on the data provided at DP2, focusing on a range of issues including errors, structure, and whether the writers' stance could be clearly identified and developed throughout the assignment. Data collected at DP 3 was hand-written and collected under exam conditions in a 3 h test, with students free to take either the 'essay' or 'report' question (1500 words).

Table 1
Learners' demographics.

Variable	Frequency
Major	
Hard	51
Soft	30
Unknown	5
L1 Background	
Mandarin	14
Cantonese	70
Other	2
MOI of Secondary School	
English Only	13
Mandarin Only	10
Cantonese Only	19
Eng-Cant	35
Eng-Mand	3
Mand-Cant	0
Eng-Mand-Cant	6
Languages other than Eng/Mandarin/Cantonese	
German	3
Japanese	1
Korean	2
French	1
Multiple	1
None	78
Time lived abroad	
Less than 1 year	3
1–2 years	4
3–4 years	1
More than 5 years	6
Never lived abroad	72

While it is possible students may produce fewer stance elements under exam conditions (DP1 and DP3) as compared to writing produced out of class (DP2), the target word counts at DP1 and DP2 were the same, and the overall normalised frequencies of all annotated items combined (per 1000 words) is actually less at DP2 (16.5 per 1000 words) than found at DP1 (20.2 per 1000 words). Moreover, the extended word count for DP3 over DP2 should allow writers to produce as many (if not more) stance features as they would be likely to produce when not under exam conditions.

4.3. Task variables

The two genres available to students are those of 'essays' and 'reports', with DP1 and 2 having two separate external secondary readings each from which to derive facts and information for the writing and for citations, with DP3 having three each. Here, our essay tasks follow Nesi and Gardner's (2012) description of essays having the social function of "developing powers of independent reasoning" where writers "demonstrate/develop the ability to construct a coherent argument and employ critical thinking skills" (pp. 37–38). The reports, in our context, have the same social function and purpose as the 'essay' prompts under Nesi and Gardner's taxonomy, with the primary difference between the two that of structure rather than argumentative function. Both essay and report prompts contain multiple stages, providing a hybrid of Nesi and Gardner's 'exposition', 'discussion', 'factorial', and 'consequential' essay types. The prompts themselves are found in Appendix A, with the accompanying type from the taxonomy provided where appropriate. For essays, both prompts follow elements of the 'Discussion'/'Factorial' essay genres alongside that of 'Exposition'. The multi-part report prompts call for elements of the 'Discussion', 'Factorial/Consequential' and 'Exposition' essay genres, meaning the 'report' prompts require a similar need for argumentative discussion with a similar concomitant need for frequent and appropriate use of stance features, despite different readings being provided for each prompt/genre.

The EAP course materials mainly emphasise the structural differences between essay and report genres, with students encouraged to develop, highlight and support their stance across the entire text in both essay and report contexts equally. The exception to this is a provision in the EAP materials and pedagogy that essays should contain more in the way of counter-arguments and rebuttals to balance the main arguments, while reports should present more in the way of recommendations based on the evidence provided by the author. The EAP curriculum distinguishes the structural differences between these genres via awareness-raising activities sequenced from the first taught session in Week 2 to the final unit of the course. For example, in week 2, students read a full essay and report and have to identify the organisation and stance features present

Table 2
Corpus description.

Corpus	Texts	Words[a]	Av. text length
DP1 – Essays	56	29,117	519.95
DP1 – Reports	31	16,114	519.81
DP2 – Essays	29	22,539	777.21
DP2 – Reports	55	45,415	825.73
DP3 – Essays	28	28,678	1024.21
DP3 – Reports	58	63,819	1100.33
Total HKU-CAES learner corpus	257	205,682	794.54

[a] Word counts include 'references' section.

in both. In later units, students conduct genre awareness tasks such as identifying appropriate section headings when specifically writing reports (these are not present in the essay activities), while the final unit of the course asks students to consider the differences between the two genres by considering the use of 'paragraphs' (essays) vs. 'sections' (reports), 'topic sentences' (essays) vs. 'headings and subheadings' (reports), and the different move structures involved in the introduction and conclusion sections of essays and reports respectively.

As shown in Table 2, there appears to be a longitudinal shift towards report tasks from essay tasks between DP1 and DP3. Following consultation with class teachers, we believe that at DP1 more students took the essay prompt as they felt more experienced in writing essays as a result of their secondary education. Over time, however, students potentially perceived the paragraphing and move structure of this text type more difficult to organise than the numbered, headed structural organisation of the report-type questions. Thus, it is likely more students took the report prompts as they felt it was the 'easier' option, although we have no firm data to confirm this.

4.4. Instruction on stance features

EAP classes were scheduled twice a week (a one-hour class followed by a two-hour class on another day). All classes followed the same EAP curriculum and materials. In units following the diagnostic writing task in week one, students are shown annotated examples of stance features (including hedges, boosters, self-mention and attitude markers) in full essay and report exemplars, with summaries outlining the importance of these features to 'successful academic writing', followed by a whole-page summary of types of supporting evidence for a stance including statistics, expert evidence (in the form of quotations or paraphrasing), common knowledge or personal accounts. Weeks six and seven (Unit 3) are entirely devoted to stance features, with exercises including identifying features of a 'successful' academic stance by comparing two different texts, one containing a high frequency of boosters, self-mention and attitude markers as representative of non-academic discourse, with the other reducing the frequency of these features in favour of hedging. Students also practice integrating counter-arguments and rebuttals to a stance, and expressing agreement and disagreement with the stance of others. Hedging is specifically treated over two whole pages of the course book with activities involving weakening overly strong claims. For boosters, learners are instructed to tone down strong claims in academic discourse, preferring that claims be both cautious and well-justified. For self-mention, the course emphasises de-personalising written production so as to maintain an academic 'tone', while for attitude markers the course emphasises justifying claims critically, rather than appealing to emotion.

Stance is also included as the joint heaviest-weighted component of the standardized assessment rubric for the program as part of the criteria '*Ability to express academic arguments*'. This criterion comprises 30% of the grade among three other criteria ('*Ability to structure an academic text*' [25%], '*Ability to write grammatically accurately with accurate use of vocabulary*' [30%] and '*Ability to cite and reference accurately*' [15%]). Within the '*Ability to express academic arguments*' criteria, the (un-weighted) subcomponents specific to stance at grade 'A' are worded as '*You can, at all times, critically justify/argue for a consistent stance in the main question*' and '*The stance is, at all times, clear and concise. There is never any confusion or ambiguity for the reader*'. While the rubric does not specifically mention the use of hedges, boosters, attitude markers or self-mention, it is therefore assumed (and also frequently made clear to students during the course) that students' appropriate production of these stance features is crucial to the raters' interpretation of whether a student has the ability to express an academic argument 'successfully', as evidenced in the grade awarded for that criterion.

4.5. Annotation and analysis

Data was converted into plain text files, then into a searchable corpus using the corpus compilation and annotation software UAMCorpustool (version 3.3h, O'Donnell, 2008). The full list of hedges, boosters, attitude markers and self-mentions annotated is found is sourced from Hyland (2005), and is included in Appendix B. Table 3 describes the number of annotated items at each DP. After annotation, the researchers examined each item to ensure the target was not only performing a metadiscoursal function, but also to identify their specific functions (if multifunctional). For example, use of the modal 'may' be used as an epistemic hedging device (e.g. 'The impact of global warming *may* be overstated') but may also be used

deontically (e.g. 'the university *may* consider taking measures to … '). Once the researchers had checked each item for its intended function, two native speakers of English checked ten texts per task type (essay vs. report) and data-point (1–3) constituting 60 texts or 23% of the total data for appropriateness of annotations (correct/incorrect). Raters checked the same texts, with rater agreement measured via Intraclass Correlation Coefficient (ICC). The ICC score was 0.738, which is considered a 'good' quality of rater agreement (Fleiss, 1981).

Raw frequencies of annotated items were converted into a normalised frequency per 1000 words in UAMCorpustool for each corpus file, with statistical analysis conducted in SPSS (v.20). For analysis of wordings unique to particular data points, UAMCorpustool was used to generate frequency wordlists which were then entered into the *log-likelihood calculator* (Rayson, http://ucrel.lancs.ac.uk/llwizard.html) to determine the log-likelihood value (*LL*) of each wording (i.e. a contingency value representing potential significant differences between target wordings and non-target wordings in two [or more] corpora). Significance values of $p < 0.01$ are given for *LL* values of greater than 6.63, $p < 0.001$ for *LL* values of greater than 10.83, and $p < 0.001$ for *LL* values of greater than 15.13. Effect size of the LL measure is determined via the *Effect Size for Log Likelihood* measure (ELL, Johnston, Berry, & Mielke, 2006), included with the Rayson calculator.

5. Results

5.1. Longitudinal variation in L2 expression of stance

Table 4 shows the longitudinal distribution of stance features across the corpus ('essays' and 'reports' combined). As non-parametric statistics are performed, the median and median absolute deviations are presented rather than the mean and standard deviations. To avoiding 'data dredging'/'data fishing', where multiple tests can lead to false significance (Young & Karr, 2011), corrected alpha values (0.0125) are used before testing with Dunn's correction applied to significant *p* values.

The data reveal a longitudinal rise in the use of hedges at later DPs from that seen at DP1, alongside a quantitative reduction of boosters, self-mention and attitude markers. The rise in hedges is presumably suggestive of an impact of instruction, given the positive emphasis placed on hedging claims throughout the course. The reduction in the frequency of boosters is in line with the rise in hedging, emphasising the need to tone down strong claims in academic discourse as frequently suggested during the course. For self-mention, the reduction of these forms is also presumably as an effect of instruction given the course's emphasis on depersonalising written production, while an emphasis on justifying claims rather than appealing to emotion is presumably responsible for the longitudinal reduction of attitude markers.

We then attempted to provide further inferential support to these claims using a series of mixed-effects linear regression models (one per stance feature). In the first instance, we generated standardized z-scores for the normalised frequencies of hedges, boosters, attitude markers and self-mention per file so as to allow for linear regression, then added the participant's demographic information (Table 1) as crossed random effects, with DP number as the repeated measures variable. DP and task type (essay/report) were added as fixed variables, with estimation performed using restricted maximum likelihood. For hedges, DP (taken as impact of instruction) was a significant predictor of the use of this device ($F_{(2, 87.94)} = 15.97, p < 0.001$), as DP was also for boosting ($F_{(2, 155.58)} = 31.68, p < 0.001$), attitude markers ($F_{(2, 83.60)} = 8.81, p < 0.001$) and self-mention ($F_{(2, 185.39)} = 18.65, p < 0.001$). Bonferroni-corrected pairwise comparison showed significant increase of hedging at DP2 over DP1 (mean diff $= 0.355$, df $= 81.38, p = 0.049$). The same type of comparison showed a significant decrease of boosting at DP3 over DP1 (mean diff $= -0.758$, df $= 222.43, p < 0.001$) and DP3 over DP2 (mean diff $= -0.495$, df $= 119.17, p < 0.001$), a significant decrease of attitude markers at DP3 over DP1 (mean diff $= -0.574$, df $= 138.28, p = 0.001$) and DP3 over DP2 (mean diff $= -0.270$, df $= 59.14, p = 0.017$), and a significant decrease of self-mention at DP3 over DP1 (mean diff $= -0.387$, df $= 368.31, p = 0.001$). Task type (essay vs. report) was a significant predictor of hedging ($F_{(1, 153.67)} = 46.62, p < 0.001$), as it was for boosting ($F_{(1, 241.90)} = 87.06, p < 0.001$) and self-mention ($F_{(1, 507.27)} = 17.37, p < 0.001$). There was also a significant interaction between hedging, task type and DP ($F_{(2, 132.74)} = 10.02, p < 0.001$) as well as for boosting ($F_{(2, 239.58)} = 7.21, p = 0.001$) and self-mention ($F_{(2, 337.15)} = 3.28, p = 0.039$), suggesting unequal variance by task type and occasion, necessitating separate investigation of these stance features across task type and DP (see section 5.2 below). For

Table 3
Annotated items per DP/Task type.

Corpus	Annotated items[a]
DP1 — Essays	H = 315, B = 230, SM = 82, AM = 40
DP1 — Reports	H = 99, B = 98, SM = 15, AM = 30
DP2 — Essays	H = 312, B = 190, SM = 10, AM = 31
DP2 — Reports	H = 335, B = 174, SM = 6, AM = 53
DP3 — Essays	H = 229, B = 161, SM = 34, AM = 26
DP3 — Reports	H = 443, B = 190, SM = 33. AM = 45
Total HKU-CAES learner corpus	H = 1733, B = 1043, SM = 180, AM = 225

[a] H= Hedges, B=Boosters, SM=Self-mentions, AM = Attitude Markers.

Table 4
Longitudinal distribution of stance markers.

Feature	Median (M)/Absolute deviation (AD) per 1000 words	Kruskal-Wallis	Pairwise Comparison (BOLD = statistically significant)
Hedges	(DP1) M = 5.12 AD = 2.44 (DP2) M = 7.25, AD = 3.33 (DP3) M = 6.56, AD = 2.24	$H(2) = 9.17, p = 0.010$	**DP3 > DP1 $t(2)$ = 26.99, p = 0.017** **DP2 > DP3 $t(2)$ = 32.04, p = 0.005** DP1 = DP2 $t(2)$ = −5.08, p = 0.660
Boosters	(DP1) M = 6.02, AD = 2.68 (DP2) M = 3.95, AD = 1.98 (DP3) M = 2.71, AD = 1.32	$H(2) = 37.65, p < 0.001$	**DP1 > DP2 $t(2)$ = 30.72, p = 0.021** **DP2 > DP3 $t(2)$ = 38.50, p = 0.002** **DP1 > DP3 $t(2)$ = 69.24, p < 0.001**
Self-Mention	(DP1) M = 0, AD = 0 (raw = 97) (DP2) M = 0, AD = 0 (raw = 16) (DP3) M = 0, AD = 0 (raw = 67)	$H(2) = 29.35, p < 0.001$	**DP1 > DP3, $t(2)$ = 30.06, p = 0.002** **DP1 > DP2, $t(2)$ = 47.69, p < 0.001** DP2 = DP3, $t(2)$ = −17.90, p = 0.139
Attitude Markers	(DP1) M = 1.21, AD = 1.21 (DP2) M = 0.99, AD = 0.99 (DP3) M = 0.24, AD = 0.24	$H(2) = 7.74, p = 0.021$	**DP1 < DP3, $t(2)$ = 26.38, p = 0.042** DP2 = DP3, $t(2)$ = 25.46, p = 0.056 DP1 = DP2, $t(2)$ = 0.92, p = 1

each of the analyses above, at no point was major, L1 background, MOI, other languages spoken or time lived abroad flagged as significant covariance predictors, suggesting these were not important random factors in the use of these devices.

In terms of the differences in wordings across DPs, Tables 5 and 6 show the distribution of the ten most frequent wordings across DPs (Table 5) before comparing the wordings specific to a particular DP (Table 6).

Tables 5 and 6 suggest longitudinal variation in the selection of a wide range of hedges, boosters, attitude markers and self-mention. For hedges, the expressions *claim, relatively, fairly,* and *in my opinion* are features of L2 production at DP1 before we see the expressions *probably, usually* and *argue* feature at DP2. *Probably* is commonly used to hedge a variety of verbs such as 'explain' or 'cause', while the habitual adverb *usually* is commonly found with verbs such as 'provided' or 'regarded'.

1) [DP2 − 0022Q6-2.txt] As a result, this *probably* explains the surge in death penalty exonerations.
2) [DP3 − 0012O2-3.txt] Classroom and real world development experiences are *usually* provided independently.

For boosters, students coming into the EAP course use a wide range of boosting expressions before narrowing this range as the course progresses. Notably, categorical statements like *never* and *sure,* common at DP1, are avoided at later DPs, while expressions assuming the readers' knowledge state such as *obvious* are also avoided as the course progresses:

3) [DP1 - 010U2.txt] If an innocent is not judge properly, they *never* have a change to be helped
4) [DP1 - 0102Z4.txt] *For sure* adopting the child penalty again could reduce the violent crime
5) [DP1 - 0092O2.txt] It is *obvious* that Hong Kong will be able to control over crimes without any death penalty.

For self-mention, students at DP1 are much more likely to insert themselves into texts. At DP1 *I* has a variety of uses, including outlining the structure of the essay/report, as a sequencing device in the middle of the writing, and to stamp personal attitude on claims or recommendations made at the end of the writing.

6) [DP1 − 0062N4.txt] *I* will first state the potential benefits of the imposition of the death penalty
7) [DP1 - 0162O2.txt] As *I* have mentioned in the beginning, Hong Kong has already been dubbed the safest city on Earth
8) [DP1 − 0072N4.txt] In conclusion, not much Hong Kong people can accept the death penalty, therefore, *I* think death penalty should not restore in Hong Kong.

My is invariably used in the phrase *in my opinion/view*:

9) [DP1 - 0182Q6.txt] *In my opinion*, the death penalty shouldn't be restored in Hong Kong because the effectiveness and efficiency of the execution on death penalty is in question.

For attitude markers, we see a reduction in authors overtly stating how the topic of discussion is *important*, with this term perhaps considered too emotional following instruction, while authors are also less likely to categorically *disagree* with the claims of others or with the topic under discussion. We also see the complete avoidance of using exclamation marks as attitude markers at later DPs:

10) [DP1 − 0032N4.txt] It is not only *important* solution, but also a useful road to deal with this problem
11) [DP1 − 0072N4.txt] In the case of Hong Kong, I *disagree* the death penalty

Table 5
Wordings of stance features across DPs.

DP1	Raw/Norm. Freq per 1000 words		DP2	Raw/Norm Freq. per 1000 words		DP3	Raw/Norm Freq. per 1000 words	
Hedges								
may	44	0.85	may	76	0.92	may	112	0.99
suggest	28	0.54	possible	34	0.41	suggested	77	0.68
often	21	0.41	usually	33	0.40	might	41	0.36
possible	18	0.35	often	32	0.39	possible	37	0.33
argue	18	0.35	might	31	0.37	usually	31	0.28
claimed	18	0.35	suggested	30	0.36	often	30	0.27
would	16	0.31	argue	29	0.35	could	29	0.26
might	15	0.29	suggest	29	0.35	about	26	0.23
fairly	15	0.29	could	20	0.24	would	20	0.18
claim	14	0.27	around	17	0.21	suggest	19	0.17
Boosters								
find	25	0.48	clear	27	0.34	believed	24	0.23
found	23	0.44	found	25	0.32	clear	22	0.21
believe	22	0.43	find	23	0.29	know	22	0.21
always	21	0.41	always	22	0.28	always	20	0.19
show	21	0.41	must	20	0.25	found	19	0.18
believed	17	0.33	believed	17	0.21	certain	17	0.16
clear	16	0.31	true	17	0.21	believe	14	0.14
shown	15	0.29	in fact	15	0.19	must	15	0.14
actually	14	0.27	believe	14	0.18	indeed	15	0.14
never	14	0.27	actually	14	0.18	show	13	0.13
Attitude Markers								
important	29	0.99	important	29	0.55	essential	19	0.31
essential	7	0.24	essential	9	0.17	admittedly	10	0.16
unfortunately	4	0.14	admittedly	7	0.13	unfortunately	6	0.1
disagree	4	0.14	inappropriate	7	0.13	expected	5	0.08
!	4	0.14	expected	3	0.06	remarkable	5	0.08
ought	3	0.1	ought	3	0.06	interesting	3	0.05
prefer	3	0.1	prefer	3	0.06	hopefully	3	0.05
importantly	3	0.1	importantly	3	0.06	fortunately	3	0.05
preferable	3	0.1	appropriate	3	0.06	inappropriate	2	0.03
agree	3	0.1	unfortunately	2	0.04	important	1	0.02
Self-Mention								
I	72	2.84	I	11	1.11	I	43	1.37
my	16	0.63	we	3	0.3	we	16	0.51
we	8	0.32	my	1	0.1	my	7	0.22
me	1	0.04	me	1	0.1	author	1	0.03
author	0	0	author	0	0	me	0	0

12) [DP1 − 0102N4.txt] Although the overall crime rate decreased by approximately 4%, the number of homicide cases increased significantly by over 120%*!*

In summary, the longitudinal variation of stance features exhibited between DP1 and DP3 appears to result in a distribution and usage more in line with the conventions of academic discourse as the result of EAP instruction.

5.2. Longitudinal variation of stance features by task type

Given the findings above that task type (essay vs. report) and the interaction between task type and DP were significant predictors of the use of hedging, boosting and self-mention devices, we then determined whether instruction has any longitudinal effect on appropriate genre-specificity of stance features. Mann-Whitney U tests were used to support our regression models in determining whether task type (essay vs. reports) was a contributor to the distributions of stance features present in both essay and reports at each DP. We also observed the wordings involved with both kinds of device at each DP.

At DP1, there were significant differences between essays and reports in the distributions of hedges (essays M = 8.07, AD = 3.7, reports M = 4.95, AD = 1.98, U = 443, t = −3.767, $p < 0.001$) and self-mention (essays M = 1.46, AD = 1.46, reports M = 0, AD = 0, U = 544, t = −3.143, $p = 0.002$). The use of the hedges *may* (LL = 24.90***, ELL = 0.00020), *argue* (LL = 15.88***, ELL = 0.00019), and *fairly* (LL = 13.23**, ELL = 0.00017) are particular to essays, while *suggest* (LL = 9.14*, ELL = 0.00009) is particular to reports. The self-mentions *I* (LL = 12.79**, ELL = 0.00009) and *my* (LL = 7.81*, ELL = 0.00010) are also particular to essays. The finding for self-mentions is primarily indicative of the differences in structuring between essays and report

Table 6
Wordings specific to a particular DP.

Word	Freq. DP1 (raw)	Freq. DP2 (raw)	Freq. DP3 (raw)	Greater Freq. in.	Log Likelihood	Effect Size (ELL)
Hedges						
Claim (lemma)	32	29	8	DP1	40.03***	0.00007
Relatively	12	13	4	DP1	11.33**	0.00003
Fairly	15	3	1	DP1	29.61***	0.00010
In my opinion	7	1	1	DP1	12.71**	0.00009
Probably	0	13	5	DP2	15.52***	0.00005
Usually	7	33	31	DP2	9.47*	0.00002
Argue (lemma)	26	43	27	DP2	11.57**	0.00002
Boosters						
Find (lemma)	50	49	30	DP1	30.78***	0.00004
Believe (lemma)	42	32	38	DP1	14.12**	0.00002
Never	14	11	7	DP1	10.04*	0.00003
Obvious	14	3	7	DP1	15.42***	0.00005
Really	12	1	8	DP1	6.82*	0.00002
Sure	7	5	1	DP1	10.53*	0.00005
Self-mentions						
I	72	11	43	DP1	84.50***	0.00012
My	16	1	7	DP1	25.29***	0.00007
We	8	3	16	DP1	7.06*	0.00002
Attitude markers						
Important	29	29	1	DP1	63.13***	0.00012
Disagree	4	1	0	DP1	9.33*	0.00043
!	4	0	0	DP1	12.12**	0.00046
Admittedly	0	7	10	DP2	8.45*	0.00003

introductions, with most examples of '*I*' being used as signposting markers and with *my* used to introduce the opinion or views of the author at the beginning of essays, as seen in this common example:

13) [DP1 − 0102N4.txt] In this essay, *I* am going to express *my* views on whether the death penalty should be restored in Hong Kong.

At DP2, there were significant differences between essays and reports for hedges (essays M = 11.39, AD = 3.67, reports M = 5.06, AD = 2.1, U = 275, t = −4.913, $p < 0.001$) and boosters (essays M = 7.25, AD = 1.21, reports M = 2.96, AD = 1.04, U = 119, t = −6.384, $p < 0.001$), with reports significantly less likely than essays to exhibit these features. The hedges *may* (LL = 55.65***, ELL = 0.00025), *argue* (LL = 37.19***, ELL = 0.00024) and *claim* (LL = 20.46***, ELL = 0.00018) and the booster *certain* (LL = 12.87**, ELL = 0.00014) are features of essays, with no particular wording specific to reports.

At DP3, there were significant differences between essays and reports for boosters (essays M = 4.55, AD = 2.01, reports M = 2.16, AD = 1.22, U = 420, z = −3.610, $p < 0.001$). The use of *indeed* (LL = 15.32, AD = 0.00011), *believe* (LL = 13.44, AD = 0.00010) and *obvious* (LL = 9.05, AD = 0.00013) are specific to essays, with no particular wording specific to reports. Notably, *indeed* appears often as a sentence-initial intensifier in later submissions, having featured in a particular exercise in Unit 3 of the course.

14) [DP3 − 0182Q6-3.txt] *Indeed*, the challenge can be adjusted by well-coordination of faculties

These results suggest a longitudinal effect on the use of stance features particular to essays, which appear to include a more frequent and wider range of such features than seen in reports. Notably, the prevalence of the boosters *indeed*, *believe* and *obvious* at DP3 showed the writers preferred to boost certain claims in their essays while perhaps making more tentative recommendations in their reports. The considerable reduction in the amount of self-mention used in essay introductions over time was also a key longitudinal trend in the data.

5.3. Stance and grade

Each assignment at DP2 and DP3 was graded for '*Ability to express academic arguments*'. At DP2, 16 texts were graded 'A', 63 texts at grade 'B', and 5 texts at grade 'C'. At DP3, 10 texts were graded 'A', 59 texts at 'B', and 15 texts at grade 'C'. No texts were graded as 'D' or 'F' at DP2 and only 1 text was graded 'D' and 'F' at DP3, and so only the data of texts graded 'A'-'C' are discussed

here. Table 7 shows the distribution of stance markers present in assignments graded 'A' to 'C' (essays and reports, DP2 and DP3 combined).

The data suggest a significant effect of the frequency of hedges and boosters on the grade assigned to a particular text, with 'A' grade texts containing more hedges and boosters than 'C' grade texts. The finding for hedging is in line with Lee and Deakin (2016) in that more successful texts contain more frequent uses of hedging, while the finding for boosters is unusual in that there is a qualitative reduction in boosters between DP1 and DP3 as the result of instruction.

Tables 8 and 9 show the wordings of stance features across and particular to a given grade. The hedges *may, might* and *could* function as modals to soften their associated main verb, while *often* is used a hedging habitual expression, and *argue* is used to hedge the claims of the author and others. Self-mention, on the other hand, is a negative predictor of grade, with students who frequently self-mention using *I* more likely to receive a 'C' grade. Given that grading decisions appear to be tied to the use of stance features, one interesting question is whether those students who received a particular grade at DP2 and who *increased* their grade (i.e. from 'C' to 'B' or from 'B' to 'A') at DP3 had achieved this increase via the use of stance features in their writing. In total, eleven students reported grade increases between DP2 and DP3. For these eleven students, we then compared the normalised frequencies of stance features present in their data between DP2 and DP3 via Mann-Whitney *U* test, with a significantly higher use of hedges reported at DP3 (U = 221, z = 2.635, p = 0.008), but not for other stance features. Therefore, for these students, increasing the frequency of hedges in their writing is likely to have been a major influence in them receiving a better grade between the writing produced at DP2 and DP3.

6. Discussion

This study has explored, in fine detail, the longitudinal impact of EAP instruction on L2 students' development of stance features present in written academic essays and reports, both in terms of the frequency and wording of such features across time, and the derivation of features constitutive of 'successful' presentations of stance as determined by EAP professionals. The results are, at the time of writing, one of the largest longitudinal corpus-based analyses of stance features present in Asian L2 EAP writing, with numerous implications for EAP and L2 instruction and assessment.

Regarding RQ1 (longitudinal variation in distribution and wording of stance features resulting from instruction), significant longitudinal differences in the use of hedges, boosters, self-mention and attitude markers are revealed in the analyses outlined in this paper. The main longitudinal impact of EAP instruction is that of a quantitative *reduction* of stance markers overall, given the course's focus on making cautious claims, depersonalising writing, and the maintenance of appropriate academic tone. It is apparent that students arriving at university are more than capable of presenting their stance on a given topic, and do so using a wide variety of available stance devices. Students entering university write with considerable energy and attitude with regard to the claims they make and the positions they take on these generic subjects (Aull, 2015), and also very keen to directly insert themselves as representative champions of said claims and position, or as a kind of 'personal assistant' directly involved in organising the text for the reader. The impact of EAP instruction, in terms of our data, is to allow students to develop a new, academic voice, gaining control over the rhetorical and linguistic aspects of academic discourse via the use of a more careful, narrower, less polarising and less personal range of expressions with which to convey their attitudes on a given topic.

In terms of the potential impact of learner-internal factors on our results, our mixed-effects models suggested that neither L1 background nor schooling was considered to be indicative of the stance features produced, and nor was the

Table 7
Stance features across assignment grade.

Feature	Median/AD per 1000 words	Kruskal-Wallis	Pairwise Comparison (Mann Whitney U, Holm-Bonferonni correction.) (BOLD = Statistically significant)
Hedges	(A) M = 7.76, AD = 2.83 (B) M = 6.31, AD = 2.66 (C) M = 4.45, AD = 0.80	*H(2) = 17.02, p < 0.001*	**A > C U = 496, z = 4.01, p < 0.001** **B > C U = 1916, z = 3.27, p = 0.001** A = B U = 2025, z = 1.94, p = 0.052
Boosters	(A) M = 4.81, AD = 2.12 (B) M = 3.38, AD = 1.61 (C) M = 1.85, AD = 0.57	*H(2) = 6.69, p = 0.035*	**A > C U = 425, z = 2.57, p = 0.010** B=C U = 1668, z = 1.88, p = 0.059 A = B U = 1920, z = 1.42, p = 0.155
Self-mention	(A) M = 0, AD = 0 (B) M = 0, AD = 0 (C) M = 0, AD = 0	*H(2) = 3.79, p = 0.150*	*N/A*
Attitude Markers	(A) M = 0.48, AD = 0.48 (B) M = 0.71, AD = 0.71 (C) M = 0.68, AD = 0.68	*H(2) = 0.14, p = 0.931*	*N/A*

Table 8
Wordings of stance features by grade.

'A' graded assignments (25,788 words)	Raw/Norm. Freq		'B' graded assignments (112,930 words)	Raw/Norm Freq.		'C' graded assignments (20,461 words)	Raw/Norm Freq.	
Hedges								
may	41	1.3	may	135	0.98	may	12	0.48
might	32	1.01	suggested	77	0.56	suggested	10	0.4
suggested	20	0.63	possible	57	0.42	usually	8	0.32
often	17	0.54	usually	51	0.37	mainly	6	0.24
argue	16	0.51	often	45	0.33	about	5	0.2
could	14	0.44	might	37	0.27	suggest	4	0.16
possible	10	0.32	suggest	37	0.27	might	3	0.12
would	8	0.25	could	33	0.24	possible	3	0.12
suggest	7	0.22	about	27	0.2	seems	3	0.12
around	7	0.22	argue	26	0.19	almost	3	0.12
Boosters								
clear	13	0.43	believed	36	0.28	always	7	0.3
always	12	0.39	clear	34	0.27	certain	7	0.3
found	9	0.3	found	31	0.24	found	4	0.17
true	7	0.23	find	25	0.2	must	4	0.17
actually	7	0.23	know	26	0.2	indeed	4	0.17
in fact	6	0.2	must	24	0.19	find	3	0.13
believed	5	0.16	always	22	0.17	show	3	0.13
must	5	0.16	believe	22	0.17	really	3	0.13
find	5	0.16	true	20	0.16	clear	2	0.08
certain	5	0.16	shown	20	0.16	true	2	0.08
Attitude Markers								
important	5	0.29	important	23	0.3	essential	5	0.31
essential	5	0.29	essential	18	0.23	inappropriate	3	0.18
admittedly	4	0.23	admittedly	12	0.16	important	2	0.12
expected	2	0.12	unfortunately	7	0.09	expected	2	0.12
importantly	2	0.12	inappropriate	6	0.08	admittedly	1	0.06
unfortunately	1	0.06	remarkable	5	0.07	hopefully	1	0.06
hopefully	1	0.06	expected	4	0.05	appropriate	1	0.06
fortunately	1	0.06	appropriate	4	0.05	interesting	1	0.06
desirable	1	0.06	hopefully	3	0.04	prefer	1	0.06
I agree	1	0.06	interesting	3	0.04	fortunately	1	0.06
Self-Mention								
I	4	1.02	I	30	1.09	I	20	2.05
my	4	1.02	we	17	0.62	we	2	0.21
me	1	0.25	my	2	0.07	my	2	0.21
we	0	0	author	1	0.04	me	0	0
author	0	0	me	0	0	author	0	0

Table 9
Stance markers wordings particular to grade of assignment (Minimum LL 6.63 p < 0.010*, 10.83, p < 0.001**, LL 15.13, p < 0.0001***).

Word	Freq. 'A' (25,788 words)	Freq. 'B' (112,930 words)	Freq. 'C' (20,461 words)	Greater Freq. in.	Log Likelihood	Effect Size (ELL)
Hedges						
May	41	135	12	'A' grade	10.84**	0.00002
Might	32	37	3	'A' grade	33.96***	0.00010
Often	17	45	0	'A' grade	19.94***	0.00006
Argue (lemma)	20	39	6	'A' grade	8.96*	0.00003
Could	14	33	2	'A' grade	7.86*	0.00003
Self-mentions						
I	4	30	20	'C' grade	21.40***	0.00007

students' major. Outside of readings and lectures for the students' major subjects - of which there are surprisingly little for freshman undergraduates to deal with in the first semester of university life - there is also little reason to expect that activities undertaken by freshman such as sports clubs or administrative societies will have a major impact on students use

of stance markers in their academic *writing*. Therefore, we assume that the primary catalyst underlying the stark change in the use of stance features is representative of the longitudinal impact of EAP instruction alone.

Regarding RQs 2 and 3 (longitudinal genre-specific variation in stance features and longitudinal variation in 'successful' stance features by grade), our findings suggest that while the frequency of stance features may be indicative of longitudinal development of an academic register as a result of instruction, the wording of such features is also crucial. For RQ2, our finding that students frequently self-insert into their essays (but not their reports) show that students do enter university with some knowledge of genre-specific differences in text organisation, with the longitudinal effect that of stripping away self-mention in favour of other, less personal means to signpost their introductions and to stamp their authority on the claims they are making. The finding regarding the longitudinal increase in boosters in essay questions (despite the quantitative drop in such features longitudinally) suggests that EAP instruction results in writers experimenting with boosting certain claims in their essays while making more tentative recommendations in their reports. For RQ3, the finding that the frequency and wording of stance features is a major indicator of the perceived 'success' of an academic text (at least in the minds of those rating them) is seen in other studies such as Lee and Deakin (2016) in terms of frequency, and Nesi and Gardner (2012), Miller et al. (2014) and Llinares and Dalton-Puffer (2015) in terms of wordings. The need to present *appropriate* (and relevant) stance in terms of the wording conventions of the academic register is crucial to the perceived success in presenting an academic argument. From our data, this appears to be the main value of the EAP course in question, with teachers serving as navigators of the academic register, aiding students in developing the necessary competence and repertoire in academic English that is a requirement for eventual success in their tertiary studies. Notably, we have shown that for students who do eventually increase the frequency of hedges while — carefully- increasing the frequency of 'appropriate' boosters, they achieve a higher degree of success in presenting their academic arguments (at least in terms of how these texts are rated by EAP professionals).

In terms of the implications for pedagogy arising from this analysis, it is clearly important for teachers to explicate to students the rhetorical and interactional effects stance markers can achieve in academic discourse. Rather than a personal take on a topic, students need to be made aware of stance as a dialogical positioning which acknowledges other views and is accessible to readers. However, instead of assigning students lists of *all* words that function as stance markers, teachers and course administrators need to focus their attention on the most frequent stance markers representative of the academic register and genres their students are attempting to write, so as to potentially decrease the amount of time taken to transition from secondary to tertiary expectations of academic discourse. Explicit instruction of the rhetorical functions of stance features may also be fruitful in raising students' awareness of achieving generic communicative functions through persuasive deployment of academic stance. In addition, the importance of teacher feedback in highlighting the (in) appropriateness of particular stance markers is also crucial, in order to unmask "these pervasive yet 'hidden' dimensions" (Lee & Deakin, 2016, p. 32) or make visible the hidden curriculum (Legg, 2016) of academic stance to the learners in their classrooms. We contend that the corpus data outlined in this study provides hard, quantifiable evidence of these hidden dimensions. The corpus data, should, therefore, be utilised by EAP teachers in lesson planning and materials preparation, as well as directly consulted by students during the writing process as part of inductive, self-guided data-driven learning of the academic register (e.g. Leńko-Szymańska & Boulton, 2015). Activities designed around the comparison of stance features across data points as well as the comparison of stance features in high vs. low rated texts would be ideal for raising awareness of register-appropriate practice — a process made all the more authentic via the use of real student submitted assignments from the course in question. Future studies are welcome to explore students' perception on the change of stance features they use, which can give a more informed investigation of longitudinal development of academic literacy and the effectiveness of EAP instruction for these stakeholders.

Due to the different tasks, conditions and prompts that the texts at each DP are produced in, the HKU-CAES corpus risks criticism for not being a controlled, experimental dataset. However, Gilquin (2016) suggests that as long as learner corpus compilers note the potential effect certain individual variables may have on the results, there is no need to attempt to control for every variable when designing learner corpora if the researcher wants to have analyse data collected in a specific environment. Rather, 'local' corpora (such as our learner corpus) "invite teachers and students into the field of learner corpus research […] resulting in learner corpora being directly useful to those, for whom, ultimately, they have been compiled" (p. 29). McEnery, Xiao, and Tono (2006) also suggest that issues of balance and representativeness when designing learner corpora should "be interpreted in relative terms" (p. 73), with "pragmatism prevailing over perfection" (Gilquin, 2016, p. 20). Given the uniform selection of text types, interlanguage varieties and authentic EAP tasks and context involved in the corpus design, we are confident that the corpus is reflective of undergraduate student production in HK during the critical first few months of exposure to EAP and the academic environment afforded by the institute in question. We are also mindful of the relatively small sample size of the corpus used when considering the size of the original cohort taking the course, with the small size due to the time and resources spent on (accurately) transcribing pen-and-paper data into electronic format. A final potential limitation of the present study is that the data collected at DP1 is not graded, and therefore we are not able to determine whether the EAP course in question constituted improvements to academic stance presentation as determined by the teachers' perception of the appropriateness and quality of such stance in their writing over time from pre- to post-instruction, but only across the revision period from the final teacher written corrective feedback at DP2 to the final test at DP3.

7. Conclusion

This study has shown that for the analysis of learner language, even small longitudinal corpora can be useful for the derivation of 'successful' student texts if performance is measured against a set of criterion used for the assessment of such success. Here, though, we agree with Park (2014) in claiming that existing criterial features used for the assessment of student performance may be unreliable "until a robust statistical relationship between given identified features and the learner's linguistic competence has been established" (p. 39). We believe our findings have made such a relationship clear. For future research, what is clearly needed is more data beyond the initial EAP period, leading into the discipline-specific language enhancement the students undergo in their sophomore years until graduation. Without this data, it is perhaps difficult to determine the overall 'effectiveness' of EAP in terms of whether such provision provides students with the means to produce an appropriate academic stance so as to ensure success in their disciplinary studies, and further studies extending the longitudinal scope of this paper are therefore deemed necessary. As suggested, further studies addressing the in-class use of such corpus data, and whether such use results in improved mastery of stance features, are also required. However, the findings of the present study have demonstrated the usefulness of short-term longitudinal corpus-based studies for the analysis of the linguistic features of academic writing, and, more importantly, have shown that such data can provide an objective, quantifiable measure of overall EAP course effectiveness in terms of student ability to develop a more register-appropriate academic stance in their writing.

Appendix A. Task prompts

Essays

(DP1/2) *"Write an academic essay answering this question: 'Should the death penalty be restored in Hong Kong?' Before its concluding paragraph, your essay should answer the following questions:*

1) *What are the main arguments for and against the death penalty?* [Discussion/Factorial]
2) *Should the death penalty be restored in Hong Kong?"* [Exposition]

(DP3) - *Write an academic essay about the following topic: 'Experiential Learning in Higher Education'. Before its concluding paragraph, your essay should:*

1) *Explain the concept of experiential learning;* [Issue]
2) *Discuss the benefits and challenges of implementing experiential learning in higher education; and* [Discussion/Factorial]
3) *Evaluate the extent to which experiential learning should be incorporated as part of the curriculum at the University of Hong Kong* [Discussion/Exposition]

Reports

(DP1/2) *"Write an academic report answering this question: 'What should be done about child labour in China?' Before its concluding paragraph, your report should answer the following questions:*

1) *What are the factors contributing to the problem of child labour in China?* [Discussion/Consequential]
2) *What should policymakers and private enterprises do (or continue doing) to solve this problem?"* [Discussion/Exposition]

(DP3) *Write an academic report about the following topic: 'Campus Sustainability'. Before its concluding section, your report should:*

1) *Identify three good practices in campus sustainability at the University of Hong Kong* [Discussion/Factorial]
2) *Discuss the potential challenges of sustainable development on campus; and* [Discussion/Factorial]
3) *Recommend and fully justify three ways to improve sustainable practices on the University of Hong Kong campus.* [Discussion/Exposition]

Appendix B. Annotated items (Hyland, 2005)

Attitude markers	Boosters	Self-mention	Hedges	
admittedly	actually	I	about	suggest
agree	always	we	almost	suggested
agrees	believe	me	apparent	suppose
agreed	believes	my	apparently	suspect
amazed	believed	our	appear	tend to
amazing	beyond doubt	mine	appeared	tends to
amazingly	certain	us	appears	typical
appropriate	certainly	the author	approximately	typically
appropriately	clear	the author's	around	uncertain
astonished	clearly	the writer	assume	uncertainly
astonishing	conclusively	the writer's	assumed	unclear
astonishingly	decidedly		certain amount	unclearly
correctly	definite		certain extent	unlikely
curious	definitely		certain level	usually
curiously	demonstrate		claim	would
desirable	demonstrated		claimed	wouldn't
desirably	demonstrates		could	broadly
disappointed	doubtless		couldn't	tended to
disappointing	establish		doubt	presumably
disappointingly	established		doubtful	suggests
disagree	evident		essentially	from this perspective
disagrees	evidently		estimate	from my perspective
disagreed	find		estimated	in my view
dramatic	finds		feel	in this view
dramatically	found		felt	in our opinion
essential	in fact		frequently	in my opinion
essentially	incontestable		from our perspective	to my knowledge
even x	incontestably		generally	fairly
expected	incontrovertible		guess	quite
expectedly	incontrovertibly		in general	rather x
fortunate	indeed		in most cases	argue
fortunately	indisputable		in most instances	argues
hopeful	indisputably		in our view	argued
hopefully	know		indicate	claims
important	known		indicated	feels
importantly	must (possibility)		largely	indicates
inappropriate	never		likely	supposed
inappropriately	no doubt		mainly	supposes
interesting	obvious		may	suspects
interestingly	obviously		maybe	postulates
prefer	of course		might	
preferable	prove		mostly	
preferably	proved		often	
preferred	proves		on the whole	
remarkable	realise		ought	
remarkably	realised		perhaps	
shocked	realises		plausible	
shocking	really		plausibly	
shockingly	show		possible	
striking	shown		possibly	
strikingly	shows		postulate	
surprised	showed		postulated	
surprising	sure		presumable	
surprisingly	surely		probable	
unbelievable	think		probably	
unbelievably	thinks		relatively	
understandable	thought		roughly	
understandably	truly		seems	
unexpected	true		should	
unexpectedly	undeniable		sometimes	
unfortunate	undeniably		somewhat	
unfortunately	undisputedly			
unusual	undoubtedly			
unusually	without doubt			
usual				

References

Aull, L. (2015). *First-year university writing: A corpus-based study with implications for pedagogy*. Springer.

Aull, L. L., & Lancaster, Z. (2014). Linguistic markers of stance in early and advanced academic writing a corpus-based comparison. *Written Communication, 31*(2), 151—183.

Biber, D. (1988). *Variation across speech and writing*. Cambridge: Cambridge University Press.

Biber, D. (2006). *University language: A corpus-based study of spoken and written registers* (vol. 23). Amsterdam: John Benjamins Publishing.

Biber, D., Conrad, S., & Reppen, R. (1998). *Corpus linguistics: Investigating language structure and use*. Cambridge, UK: Cambridge University Press.

Bruce, E., & Hamp-Lyons, L. (2015). Opposing tensions of local and international standards for EAP writing programmes: Who are we assessing for? *Journal of English for Academic Purposes, 18*, 64—77. http://dx.doi.org/10.1016/j.jeap.2015.03.003.

Castello, E., Ackerly, K., & Cocceta, F. (Eds.). (2016). *Studies in learner corpus linguistics: Research and applications for foreign language teaching and development*. Berlin: Peter Lang.

Chandrasegaran, A. (2013). The effect of a socio-cognitive approach to teaching writing on stance support moves and topicality in students' expository essays. *Linguistics and Education, 24*(2), 101—111. http://dx.doi.org/10.1016/j.linged.2012.12.00.

Chang, P. (2010). *Taking an effective authorial stance in academic writing: Inductive learning for second language writers using a stance corpus*. Unpublished doctoral dissertation. The University of Michigan.

Crosthwaite, P. (2016). A longitudinal multidimensional analysis of EAP writing: Determining EAP course effectiveness. *Journal of English for Academic Purposes, 22*, 166—178. https://doi.org/10.1016/j.jeap.2016.04.005.

Evans, S., & Morrison, B. (2011). The first term at university: Implications for EAP. *English Language Teaching Journal, 65*(4), 387—397.

Fleiss, J. (1981). *Statistical methods for rates and proportions* (2nd ed.). New York: Wiley.

Gabrielatos, C. (2005). Corpora and language teaching: Just a fling or wedding bells? *TESL-EJ, 8*(4), n4.

Geertzen, J., Alexopoulou, T., & Korhonen, A. (2013). Automatic linguistic annotation of large scale L2 databases: The EF-Cambridge Open Language Database (EFCAMDAT). In *Proceedings of the 31st second language research forum*. Somerville, MA: Cascadilla Proceedings Project.

Gilquin, G. (2016). From design to collection of learner corpora. In S. Granger, G. Gilquin, & F. Meunier (Eds.), *The Cambridge handbook of learner corpus research* (pp. 9—34). Cambridge: Cambridge University Press.

Granger, S., Gilquin, G., & Meunier, F. (2016). Introduction: Learner corpus research — past, present and future. In S. Granger, G. Gilquin, & F. Meunier (Eds.), *The Cambridge handbook of learner corpus research* (pp. 1—7). Cambridge: Cambridge University Press.

Hasko, V., & Meunier, F. (2013). Capturing L2 development through learner corpus analysis. *Modern Language Journal, 97*(S1). Supplement 2013.

Hinkel, E. (2005). Hedging, inflating, and persuading in L2 academic writing. *Applied Language Learning, 15*(1/2), 29.

Hunston, S., & Thompson, G. (Eds.). (2000). *Evaluation in text: Authorial stance and the construction of discourse: Authorial stance and the construction of discourse*. UK: Oxford University Press.

Hyland, K. (2000). *Disciplinary discourses: Social interactions in academic writing*. London, UK: Longman.

Hyland, K. (2002). Options of identity in academic writing. *ELT Journal, 56*(4), 351—358.

Hyland, K. (2005). *Metadiscourse*. New York: John Wiley & Sons, Inc.

Hyland, K. (2016). Writing with attitude: Conveying a stance in academic texts. In E. Hinkel (Ed.), *Teaching English grammar to speakers of other languages*. New York: Routledge.

Hyland, K., & Tse, P. (2004). Metadiscourse in academic writing: A reappraisal. *Applied Linguistics, 25*(2), 156—177.

Ishikawa, S. (2013). The ICNALE and sophisticated contrastive interlanguage analysis of Asian learners of English. In S. Ishikawa (Ed.), *Learner corpus studies in Asia and the World*. Kobe, Japan: Kobe University, 191—118.

Jiang, F. K. (2015). Nominal stance construction in L1 and L2 students' writing. *Journal of English for Academic Purposes, 20*, 90—102. http://dx.doi.org/10.1016/j.jeap.2015.07.002.

Jiang, F. K., & Hyland, K. (2016). Nouns and academic interactions: A neglected feature of metadiscourse. *Applied Linguistics, 2016*, 1—25. http://dx.doi.org/10.1093/applin/amw023.

Johnston, J. E., Berry, K. J., & Mielke, P. W. (2006). Measures of effect size for chi-squared and likelihood-ratio goodness-of-fit tests. *Perceptual and Motor Skills, 103*(2), 412—414.

Lancaster, Z. (2016). Expressing stance in undergraduate writing: Discipline-specific and general qualities. *Journal of English for Academic Purposes, 23*, 16—30.

Lee, J. J., & Deakin, L. (2016). Interactions in L1 and L2 undergraduate student writing: Interactional metadiscourse in successful and less-successful argumentative essays. *Journal of Second Language Writing, 33*, 21—34.

Legg, M. (2016). *An exploration of the voices of a new university curriculum in Hong Kong: Implications for the teaching of English for academic purposes*. Unpublished doctoral thesis. Sydney, Australia: Macquarie University.

Leńko-Szymańska, A., & Boulton, A. (Eds.). (2015). *Multiple affordances of language corpora for data-driven learning*. Amsterdam: John Benjamins.

Llinares, A., & Dalton-Puffer, C. (2015). The role of different tasks in CLIL students' use of evaluative language. *System, 54*, 69—79.

Martin, J. R., & White, P. R. (2005). *The language of evaluation* (2nd. edition). London: Palgrave Macmillan.

McEnery, T., Xiao, R., & Tono, Y. (2006). *Corpus-based language studies: An advanced resource book*. London: Routledge.

Meunier, F. (2016). Developmental patterns in learner corpora. In S. Granger, G. Gilquin, & F. Meunier (Eds.), *The Cambridge handbook of learner corpus research* (pp. 379—400). Cambridge: Cambridge University Press.

Meunier, F., & Littré, D. (2013). Tracking learners' progress: Adopting a dual 'Corpus-cum-experimental data' approach. *Modern Language Journal, 97*(1), 61—76. http://dx.doi.org/10.1111/j.1540-4781.2012.01424.x.

Miller, R. T., Mitchell, T. D., & Pessoa, S. (2014). Valued voices: Students' use of engagement in argumentative history writing. *Linguistics and Education, 28*, 107—120.

Nesi, H., & Gardner, S. (2012). *Genres across the disciplines: Student writing in higher education*. Cambridge: Cambridge University Press.

Nicholls, D. (2003, March). The Cambridge learner corpus: Error coding and analysis for lexicography and ELT. In *Proceedings of the corpus linguistics 2003 conference* (Vol. 16, pp. 572—581).

O'Donnell, M. (2008). The UAM CorpusTool: Software for corpus annotation and exploration. In *Proceedings of the XXVI Congreso de AESLA, Almeria, Spain* (pp. 3—5).

Park, K. (2014). Corpora and language assessment: The state of the art. *Language Assessment Quarterly, 11*(1), 27—44. http://dx.doi.org/10.1080/15434303.2013.872647.

Qin, J. (2014). Use of formulaic bundles by non-native English graduate writers and published authors in applied linguistics. *System, 42*, 220—231.

Sinclair, J. (2004). *How to use corpora in language teaching*. Amsterdam, the Netherlands: John Benjamins.

Tribble, C., & Wingate, U. (2013). From text to corpus—A genre-based approach to academic literacy instruction. *System, 41*(2), 307—321.

Wingate, U. (2012). 'Argument!': Helping students understand what essay writing is about. *Journal of English for Academic Purposes, 11*, 145—154. http://dx.doi.org/10.1016/j.jeap.2011.11.001.

Young, S. S., & Karr, A. (2011). Deming, data and observational studies. *Significance, 8*(3), 116—120.

Zhao, C. G. (2013). Measuring authorial voice strength in L2 argumentative writing: The development and validation of an analytic rubric. *Language Testing, 30*(2), 201—230.