APPRAISAL RESOURCES IN L1 AND L2 ARGUMENTATIVE ESSAYS: A CONTRASTIVE LEARNER CORPUS-INFORMED STUDY OF EVALUATIVE STANCE

SUET LING LAM

UNIVERSITY OF HONG KONG

PETER CROSTHWAITE

UNIVERSITY OF QUEENSLAND

ABSTRACT

Martin and White's (2005) *appraisal* theory explores the ways interpersonal meanings are expressed when researching evaluative aspects of language use. Despite numerous discourse analyses using *appraisal*, there is little research comparing *appraisal* resources deployed by L1 and L2 English writers in discourse produced under the same conditions and on the same task. 60 argumentative essays across two writing prompts were collected from a larger corpus of Asian L2 English writing, to which the present study applied the *appraisal* model as part of a Contrastive Interlanguage Analysis (Granger, 1996, 2015) comparing the evaluative stance construed in L1 and L2 English texts. The findings show significant variation in the evaluative resources used in L1 and L2 essays. L1 English writers demonstrated a consistent reliance on *engagement* resources in general and showed a heavier dependence on these devices than L2 writers in their essays. In contrast, Hong Kong L2 English writers used a significantly higher frequency of *negative attitude* resources than L1 English writers. These contrastive corpus-informed results offer further evidence of differences in L1/L2 written evaluative stance that educators may take into account during writing instruction.

CONTACT

Peter Crosthwaite, School of Languages and Cultures, Faculty of Humanities and Social Sciences, University of Queensland, Brisbane St Lucia, QLD 4072. p.cros@uq.edu.au

ORCID

0000-0002-1482-8381

Appraisal resources in L1 and L2 argumentative essays

Suet Ling Lam   **Peter Crosthwaite**
University of Hong Kong   *University of Queensland*

1.   Introduction — Evaluation in writing research

Argumentative writing is viewed as a dynamic use of language representing real-world experiences while acknowledging, constructing and negotiating social relations (Hyland, 2005). Writers' interpersonal language use plays a crucial role in persuasive writing, providing the linguistic means for writers to introduce their authorial voice and engage with diverse viewpoints, in order to build a convincing argument and establish solidarity with readers. In particular, writing research has long attended to the evaluative aspect of interpersonal language resources. Hunston and Thompson (2000) define evaluation as "the expression of the speaker or writer's attitude or stance towards, viewpoint on, or feelings about the entities or propositions that he or she is talking about" (p. 5). The primary concern of evaluation is why, when, how and what writers evaluate through the syntactic and lexical choices they make. Evaluation is used to perform three non-exclusive functions; to express the writer's opinion, reflecting the value system of both the writer and their community; to construct and maintain relations between the writer and readers; and to organise discourse (Hunston and Thompson, 2000).

There is a large body of research looking into evaluation in writing via discourse analysis. Attention has been paid to the argument function of evaluation focusing on the notion of genre (e.g., Swales, 1990; Paltridge, 1997), while evaluative disjuncts have been argued to have a role in performing interpersonal and textual functions in academic texts (Thompson and Zhou, 2000). Evaluation is also said to play a role in the construal of ideology and in organising a persuasive text (Hunston, 2000). Other studies attending to the interpersonal dimension of writing under the umbrella term 'evaluation' include models of stance (e.g., Biber and Finegan, 1988), metadiscourse (e.g., Hyland and Tse, 2004) and *appraisal*[1] (Martin and White, 2005). In particular, Martin and White's *appraisal* model is a comprehensive description of the resources available in English to construe interpersonal meanings in a text, notably useful in analysing the construction of evaluative stance in argumentative essays, as argumentation entails complex linguistic resources.

This study is concerned with evaluation in first language (L1) and second language (L2) English essays, using Martin and White's *appraisal* system as the framework for evaluation. While a considerable amount of research on the *appraisal* system has taken a

---

1   Labels for components of the *appraisal* system are given in italics according to the conventions of Systemic Functional Linguistics.

contrastive discourse analysis approach to examine interpersonal language, much of this work has been done using conversation analytic methods on very small datasets, or has focused on published research articles (Bruce, 2014). Corpus researchers (e.g., Flowerdew, 1998; Granger, 2002, 2004; Sinclair, 2004) have long advocated the incorporation of corpus linguistic techniques to accompany the methods of analysing discourse, and research of this nature is beginning to enter the literature, such as Hood (2005) and, most recently, Geng and Wharton (2016). However, these studies have compared published research writing with L2 essays (Hood) or L1 writing from two different languages (Geng and Wharton), and there is still a need to conduct a study comparing L1 and L2 English writing produced under the same conditions on the same task(s) — a gap this paper intends to fill. This paper first outlines the *appraisal* framework for the reader, before discussing a corpus-informed approach to the analysis of evaluation under this framework.
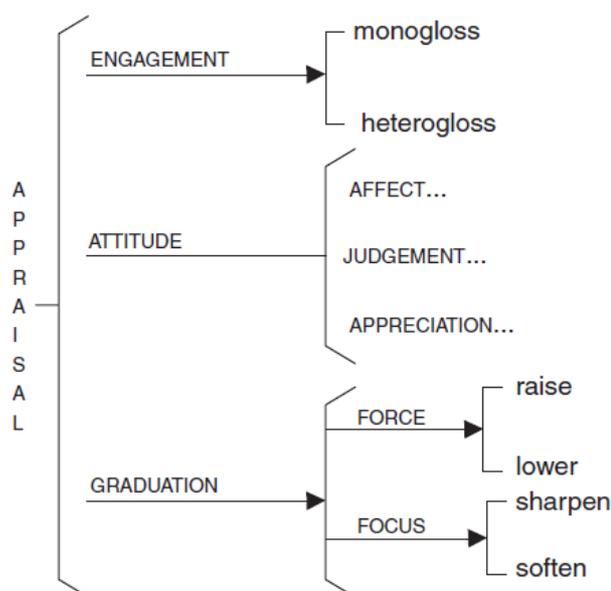

Figure 1: An overview of Martin and White's appraisal system (Martin and White, 2005, p. 38)

2. Appraisal

The *appraisal* system and its subcomponents involve the major discourse semantic resources construing the interpersonal meanings of a text across three interacting domains: *attitude, engagement* and *graduation* (see Figure 1).

2.1. Attitude

Attitude is concerned with expression of emotions (*affect*), judgement of behaviour (*judgement*) and evaluation of entities (*appreciation*). Evaluation can be directly inscribed

in discourse or invoked through various means, such as lexical metaphors that may evoke the writer's attitude, as well as vocabulary items which provoke an attitude and so are more sensitive to co-text and reading position for interpretation (Martin and White, 2005). *Affect* looks at resources for construing emotional responses, which can be positive or negative. *Affect* is sub-divided into *authorial evaluation* when the author is the one ascribing affect and *non-authorial evaluation* when a participant is the one evaluating affect. *Judgement* deals with resources for evaluating behaviour according to social esteem and social sanction. It is concerned with admiration or criticism of personal behaviour. In addition, judgement of behaviour or character can be made on moral, legal and ethical grounds. *Appreciation* is concerned with resources for construing value of things, including abstract phenomena and physical objects. Entities can be evaluated with regard to reactions they provoke, their compositional qualities and their worthiness.

2.2.   Engagement

Engagement deals with linguistic resources of intersubjective positioning, introducing the authorial voice and engaging alternative voices and positions. It provides means for writers and readers to negotiate relationships of alignment and disalignment. Drawing on Bakhtin's (1986) notion of dialogic space, utterances which do not engage other voices or positions are *monoglossic* and those which open up space for alternatives are *heteroglossic*. Heteroglossic resources are divided into two categories: dialogic expansion and dialogic contraction. Dialogic expansion allows alternative positions and voices by opening up dialogic space. There are two types of dialogic expansive resources: *entertain* and *attribute*. Words and phrases which open up space for a range of possibilities are labelled *entertain*. In the sub-category of *attribute*, other sources which are explicitly acknowledged are labelled *acknowledge* while *distance* occurs when a writer distances themself from another viewpoint through reporting verbs.

Dialogic contraction, on the other hand, limits the space for dialogic alternatives by rejecting another position using *disclaim* resources. These resources are divided into two types: *deny* and *counter*. *Deny* occurs when a writer explicitly denies another's proposition via negation. *Counter* happens when an author responds to another viewpoint with a counterargument through conjunctions and connectives. In the sub-category of *proclaim*, there are four types of resources namely *concur*, *pronounce*, *endorse* and *justify*. *Concur* is ascribed when a writer overtly introduces their positive alignment with a viewpoint. It could be either via affirm or concede. Expressions which state an author's proposition clearly are coded *pronounce*. Propositions which are formulated by citing external sources are labelled *endorse*. Expressions which justify a proposition with a reason are coded *justify*. Figure 2 gives an overview of *heteroglossic* resources in the *engagement* system.
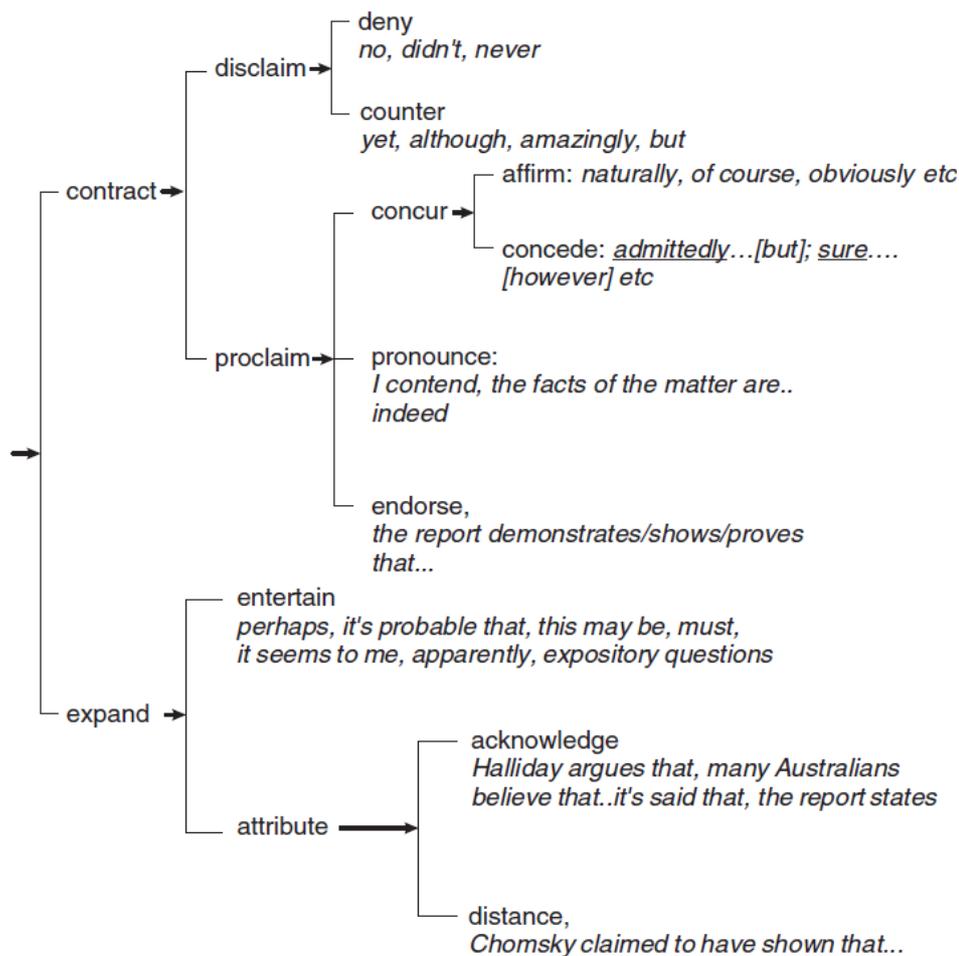
Figure 2: *Heteroglossic resources in the engagement system (Martin and White, 2005, p. 134)*

2.3. Graduation

Graduation focuses on resources which vary the strength of evaluation in a text in terms of *force* and *focus* (see Figure 3). *Force* attends to the gradability of experiential attitudinal meanings. It covers resources which *upscale* and *downscale* qualities and processes through intensification and quantification. *Focus* functions to enhance or weaken the degree of positivity or negativity. It involves sharpening and softening of boundaries around categorical meanings (Hood, 2006) which are not scalable.
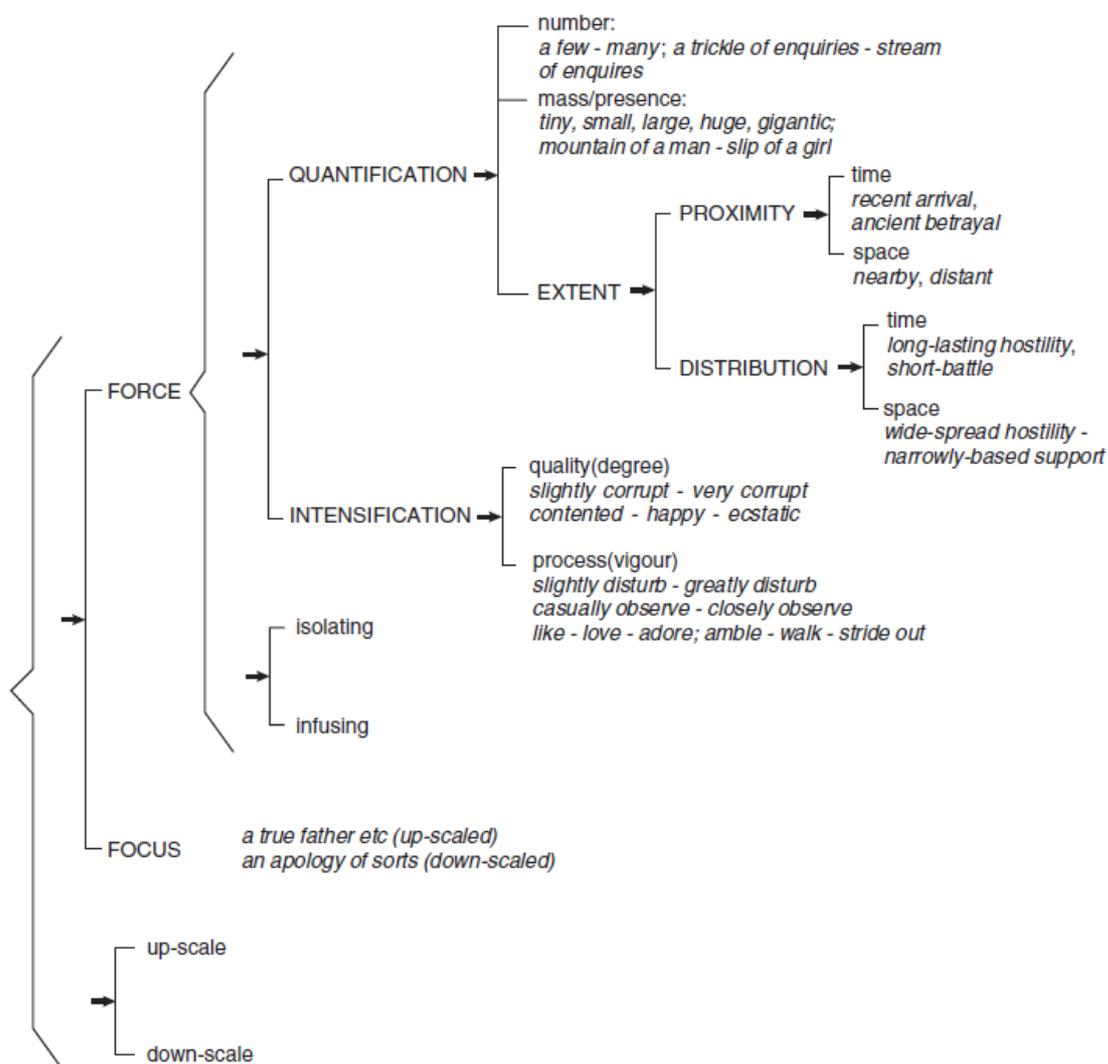
Figure 3: *The graduation system (Martin and White, 2005, p. 154)*

3.   Applications of *appraisal* theory for persuasive texts

Appraisal theory has been commonly used as a framework to research evaluative stance in both L1 and L2 writing. Specific to L2 writing, Hood (2005) explored how undergraduate students learning English as a second language construct *attitude* in their evaluative stance in the introductory sections of their research papers. Here, *appreciation* was the dominant form of *attitude* in both student texts and equivalent professional writing. However, professional writers tended to evaluate through *appreciation* while student writers also evaluated through emotional responses (*affect*) and ethical concerns (*judgement*), suggesting L2 writers construct a more personalised expression than seen in professional writing. Unlike professional writers, student texts also show limited use of *graduation* and

often ambiguously encoded positive *appreciation*. Hood concluded that students need to have an access to a range of linguistic resources to express *attitude* besides that of recognising the move structure of a genre, as commonly emphasised in many English for Academic Purposes (EAP) pedagogical writing materials.

The use of *engagement* in L2 writing has received relatively little attention compared to that of *attitude*. In a study on Chinese L2 English writers, Coffin and Hewings (2004) studied the use of *engagement* resources, particularly *hearsay* and *pronounce* under *heterogloss*, in high- and low-scoring academic argumentative essays written for the International English Language Testing System (IELTS) examination. Hearsay involves reporting on an utterance where the speaker or the writer is not specified, while *pronounce* refers to authorial emphasis or explicit authorial interventions (Martin and White, 2005). The authors found that high-scoring texts deployed a wider range of *hearsay* resources, with *hearsay* working to minimise subjectivity and open up alternative voices for negotiation between the writer and readers, while *pronounce* increases subjectivity while emphasising broader community views. Chinese writers generally, however, were more direct in making claims via *pronounce* — a practice often discouraged in academic writing, because of a limited repertoire of interpersonal resources within their interlanguage. Overuse of such resources weakens authorial voice and thus decreases the persuasiveness of any claims. In another representative study, Swain (2007) analysed English as a Second Language discursive writing using the *engagement* system to examine how authors engage with topics and readers. Swain concluded that effective use of *engagement* resources works to increase perceived persuasiveness in L2 writing, and that achieving a balance between dialogically expansive and contractive resources adds persuasiveness to textual voice. The most recent corpus-based attempt to look at *engagement* is Geng and Wharton (2016) who compared L1 Mandarin and L1 English doctoral thesis discussions sections, finding no significant differences in the employment of *engagement* resources between the two L1s and negating L1 transfer as a factor when considering the use of such resources in L2 English texts written by L1 Mandarin speakers. This suggests L2-specific difficulties recognising and formulating evaluative stance in English writing is the main concern for L2 writers.

The system of *graduation* has been a popular focus of research on tertiary L2 production. One such example is Hood's (2006) study of academic research writing in EAP, focusing on the prosodic patterning of interpersonal meanings and how they function with reference to *graduation*. Specifically, terms of quantity (amount) or quality (extent) were often found to heighten attitudinal meanings. Hood recommended that novice L2 writers need particular help with managing these linguistic resources so as to more firmly position their evaluative stance in a text. Derewianka (2007) suggests that L2 learners should extend their repertoire of *graduation* resources at different proficiencies, initially using intensifying pre-modifiers and highly graduative lexical items at first, and

then deploying resources of quantification at intermediate levels, before finally using both *force* and *focus* to strategically align with readers.

However, as mentioned, previous studies using the *appraisal* model have mainly focused on the application of one or two categories of the *appraisal* system across a small number of texts or have not compared L1 and L2 writers working on the same tasks under the same conditions. Adopting a corpus-informed approach to *appraisal* for the present study is therefore suggested, as discussed in the next section.

4.   Corpus linguistics and discourse analysis

Corpus linguistics is a research methodology relying on collections of searchable electronic language data for linguistic analysis. Learner corpus research uses learner data and aims to provide improved descriptions of language produced by second or foreign language learners, with such analysis often used to improve foreign language teaching (Granger, 2002). A large amount of research on L2 data has followed a Contrastive Interlanguage Analysis (CIA) as part of an Integrated Contrastive Model, combining traditional contrastive analysis and CIA via corpus analysis, with the corpus approach serving to complement both methodologies (Gilquin, 2001; Granger, 1996, 2015). Studies under this framework compare and contrast professionally written or L1 data as a reference variety with that of data from novice or L2 writers to discover errors, over- and underuses of target linguistic features in the learner variety against that of the reference variety.

The integration of (learner) corpus linguistics and CIA as a methodological tool accompanying traditional discourse analysis has been gaining popularity due to the significantly increased amount of data that such a methodology affords for analysis, with accompanying improvements to the representativeness of the findings. A number of recent corpus-based discourse analyses have arisen from a systemic functional linguistics perspective (e.g., Flowerdew, 2004; Green *et al.*, 2000; Hyland and Milton, 1997). Using a CIA methodology, Hyland (1994) and Hyland and Milton (1997) compared L2 English argumentative essays by L1 Cantonese speakers against those of L1 English secondary school students to examine the expression of doubt and certainty in both. The study shows significant differences between the ways L1 and L2 writers express their attitude towards a proposition, with L2 writers relying on modal hedging and boosting devices (e.g., 'could', 'should', 'would') because of the disproportionate attention they have received in L2 pedagogical writing materials. Other representative studies on Asian L2 English data include Hong and Cao (2014), where CIA was used to compare young (grade 10) Chinese, Spanish and Polish writers' use of interaction in argumentation and description texts using Hyland and Tse's (2004) taxonomy of interactive/interactional features. Young writers as a whole restricted hedging interaction to primarily modal verbs (e.g., 'would', 'could', etc.), and noted L1 effects in the use of boosting devices (e.g.,

'very', 'really', etc.) between Chinese and Spanish groups, as well as effects of text type. Crosthwaite, Cheung and Jiang (2017) used CIA to compare novice and professional evaluation in comparable corpora of dentistry research reports, finding that novice writers as a whole are more likely than professional writers to boost their claims, refer to themselves in their texts, and use a wide range of non-disciplinary specific attitude markers when evaluating their or others' findings.

Corpus-based studies have applied the *appraisal* framework to encode evaluative lexis in both professional and learner written corpora to analyse 'problem-solution' patterns in report writing (Flowerdew, 2004). Comparing corpora of novice and professionally-written technical recommendation-type reports, the concepts of inscribed and evoked in the *appraisal* theory were applied to the classification of keywords to reflect the discourse characteristics of the two corpora. Evoked lexis was often found in the 'problem' element (of the problem-solution pattern) of the professional corpus alongside inscribed lexis in the 'solution' element. In contrast, inscribed lexis is more commonly used for the 'problem' element in the student corpus. In another corpus-based study using the *appraisal* framework, Ye (2016) built a corpus of newspaper reports from three Eastern (Chinese) and three Western American newspapers focusing on differences in their portrayals of Vladimir Putin during the Ukraine crisis. The findings suggested a marked disparity between Eastern and Western *appraisal*s of Putin, noting that despite measured acknowledgement of the inappropriateness of Putin's actions in Ukraine in the reports from both regions, the Chinese media were found to be marginally more favourable towards Putin than their American counterparts. Ye suggests that such findings highlight media heavily influenced by political, social and historical factors relating to its situational context.

However, there is still no study in the literature concerning *appraisal* and the production of L2 English learners with native L1 writers produced under the same conditions and the same task. While this may evoke accusations of the comparative fallacy (Bley-Vroman, 1983; Granger, 2015), Neff-Van Aertselaer (2016) suggests that without a comparable investigation of rhetorical practices associated with the target variety, it is 'difficult to see how L1 linguistic and rhetorical characteristics associated with discourse conventions might be influencing the L2 writing' (p.276). Aertselar also notes that the findings of corpus-based discourse studies may have implications for the teaching and testing of L2 writing, suggesting that increased data using established discourse frameworks (such as *appraisal*) and studies on L2 instructional materials containing such data can highlight their practical relevance for language teaching. Given these concerns, a CIA of *appraisal* in L1 and L2 production would be timely and advantageous.

5.   The study

This study compares the features of evaluation present in argumentative essays produced by L1 writers and learner L2 English writers from L1 Cantonese (Hong Kong) backgrounds. These issues are examined through the linguistic lens of Martin and White's (2005) *appraisal* theory alongside the incorporation of a CIA methodology, with the latter enriching the primarily qualitative evidence collected under the former.

The study therefore addresses the following research question: Do L1 writers and L2 English learner writers from L1 Cantonese backgrounds vary in employing *appraisal* resources in argumentative essays produced under the same conditions?

6.   Method

6.1.   Corpus sample

The corpus used for analysis is the written version of International Corpus Network of Asian Learners of English (ICNALE) (Ishikawa, 2011, 2013), specifically the L1 English data (henceforth NES[2]) and the 'B1-2' level L2 English data from L1 Cantonese speakers sourced from Hong Kong (henceforth HKL2E). ICNALE was specifically chosen for this research as it contains L1 and L2 argumentative essays produced under the same conditions and on the same tasks, facilitating direct comparison.

ICNALE 'B1-2' level as defined by the ICNALE creators is intended to be equivalent to an 'upper' B1 level of the Common European Framework reference descriptors (Council for Europe, 2001), or around an IELTS score of 5-5.5. Unfortunately, discrepancies in how the ICNALE L2 proficiency levels were sourced across different students and L1 backgrounds has resulted in potential non-equivalence with these standardised descriptors, despite having the same labels. This leads Crosthwaite (2016) to label 'B1-2' level as 'Intermediate' for the purposes of studies using data from that level, and this is also the approach taken in this study. The ICNALE corpus is tightly controlled for CIA type analyses, with data comprised of the same genre, topics and conditions for writing the same for all participants. This allows for maximum comparability with high reliability for L1–L2 CIA across the subcorpora.

All ICNALE essays are 200 to 300 words (±10%) comprised of only two topics ('part-time job' and 'smoking ban'), written under timed conditions (40 minutes per essay) using word processors (with spell checking allowed) and without the use of dictionaries. The prompts for the two topics are shown below:

---

2   We use the label 'NES' as taken from ICNALE itself, without further exploration of the concept of what a 'native' English speaker entails. The data is sourced by L1 speakers of English from the US, UK, Canada, Australia, and New Zealand, although information as to the exact makeup of this cohort is unavailable.

'Do you agree or disagree with this statement? Use reasons and specific details to support your claim.'

[Part-time job]: 'It is important for college students to have a part time job.'

[Smoking ban]: 'Smoking should be completely banned at all the restaurants in the country.'

Data at levels at 'intermediate' level were chosen for the analysis, as the full range of evaluative resources is generally found in the texts of more advanced writers. As all annotation is conducted manually on an exhaustive, top-down set of features per file, a small randomly sampled subset of the NES and HKL2E ICNALE corpora was analysed for the study, consisting of 60 samples, 20 from the reference NES corpus (10 from each topic) and 40 (20 from each topic) from the HKL2E corpus. While the L1 and L2 corpora are of different sizes, the word count of individual samples is similar, and so the conditions for each type of *appraisal* resource to be present in both L1 and L2 texts have been met. We have also normalised frequency counts of *appraisal* resources to instances per 1,000 tokens so as to also ensure comparability across corpora. The word counts and number of texts analysed are shown in Table 1.

Task	L1		L2	
	Words	Texts	Words	Texts
Part-time job	2349	10	4677	20
Smoking ban	2223	10	4445	20
Total	4572	20	9122	40
Total corpus size = 13694 (n = 60)				

Table 1: *Word counts and number of texts in the corpus*

6.2. Annotation

All texts were coded with reference to the *appraisal* framework: *attitude*, *engagement* and *graduation* using UAM (Universidad Autónoma de Madrid) CorpusTool, version 3.3f (O'Donnell, 2008). UAM CorpusTool is a piece of software for the annotation of text corpora. It allows segments to be annotated with more than one feature, which is suitable for dealing with the often overlapping categories of the *appraisal* system. All data were coded by the first author (a bilingual Cantonese/English speaker). The coding for all texts was then double-checked by the second author (a native English speaker). A random selection of 12 coded whole-text samples (20% of the data) were checked for accuracy of

coding by two raters (both native speakers of English and both postgraduate students of Applied Linguistics) to ensure the reliability of the coding scheme and minimize subjectivity in coding. They went over each coded segment to decide if it was 'correct' or 'incorrect'. 578 coded segments were covered in total. The Intraclass Correlation Coefficient for this analysis (2 way random, general consistency, k=2) was .712, which is considered a 'good' standard of agreement (Cicchetti, 1994).

There is a compromise between the comprehensiveness of the *appraisal* system and a corpus-informed methodological approach. The extensiveness of the framework was reduced at the expense of the amount of data analysed in the study, with the researchers taking a long time to encode evaluative instances in a given text manually. Therefore, the *appraisal* framework was simplified, focusing on the three broad dimensions (*attitude*, *engagement*, *graduation*) but with a reduced set of sub-categories within each domain, in that sub-categories which were encoded only one or two times in the trial stage of annotation were not used in the final study.

The simplified framework used in the final study is shown below (see Figure 4). A summary table of the definitions and examples of the *appraisal* resources of the simplified framework used in the study is shown in the Appendix.

7. Results

7.1.   *The use of* appraisal *resources by L1 English writers and Hong Kong L2 English writers*

To determine whether language proficiency (L1 native / L2 intermediate), task type (essay / report) or the interaction of both factors was responsible for the variance found across three main *appraisal* resources (*attitude, engagement, graduation*) in the corpus, the normalised frequencies of these resources were first converted into standardized z-scores to ensure a normal distribution, before MANOVA was performed.

The MANOVA results suggest a significant effect of L1 background ($F(3, 54)=3.948$, $p=.013$, partial $\eta^2=.180$) on the use of these resources, as well as a significant effect of task type ($F(3, 54)=3.000$, $p=.038$, partial $\eta^2=.143$). There was not, however, a significant interaction between L1 background and task type on the use of *appraisal* resources ($F(3, 54)=2.470$, $p=.072$, partial $\eta^2=.121$). This rules out L1/task by occasion variance, suggesting instead that each fixed factor made its own contribution to the variance in the results. This result necessitates that separate analyses be made of L1/L2 differences in *appraisal* use across both task types.
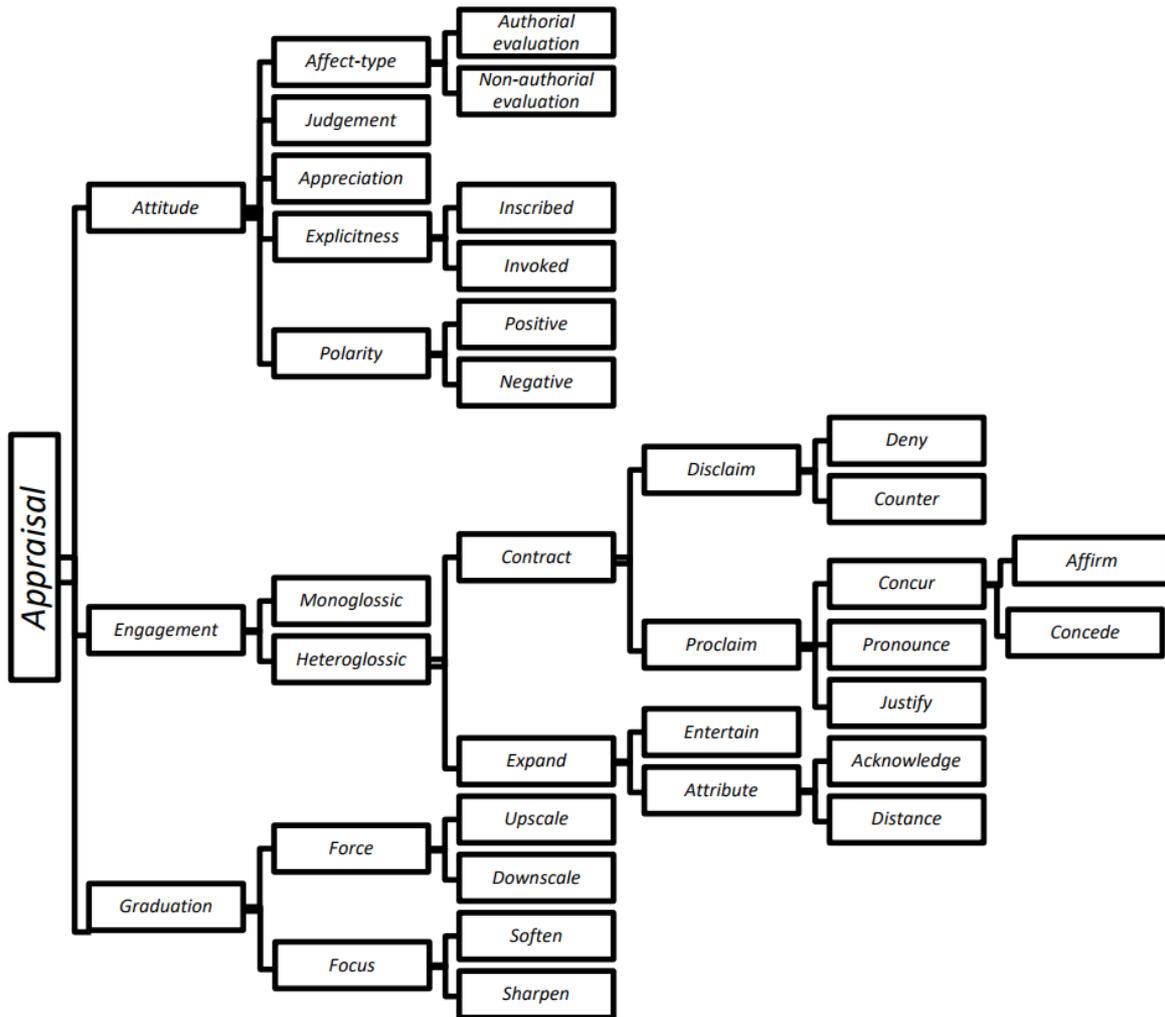
Figure 4: *The simplified appraisal framework (adapted from Martin and White, 2005, p. 38)*

Table 2 shows the raw and median normalised frequencies of the three main *appraisal* categories alongside the median absolute deviation which provides a measure of individual variation. Median values rather than means are shown as the data are generally non-normally distributed (following significant Shapiro-Wilks tests), with an adjusted alpha value of 0.00833 to correct for multiple tests for the Mann-Whitney U comparisons (*n*=6).

Both NES and HKL2E groups made use of all three domains of interpersonal resources in expressing their evaluative stance in these argumentative essays. Both L1 and L2 writers tended to use *engagement* resources the most, reflecting frequent intersubjective positioning when introducing authorial voice and engaging alternative voices and positions. The number of instances of *graduation* was slightly higher than *attitude* in essays produced by both language groups, with writers frequently intensifying semantic meanings (or diminishing their force) when making claims. L1 writers were

found to use a significantly higher frequency of devices for *engagement* than L2 writers for both writing tasks, and *graduation* devices for the 'part-time job' task. L2 writers tended to utilise a more even distribution of the main three types of evaluative resources in their writing.

	L1			L2			Mann–Whitney *U* test			
	f	*f* per 1,000 words	median absolute deviation	*f*	*f* per 1,000 words	median absolute deviation	U	*z*	*p*	Pearson's *r*
Attitude										
Part-time job	114	44.65	5.68	251	44.38	9.07	113	-.572	.588	-.104
Smoking ban	88	36.11	8.43	208	41.52	10.96	117	.748	.474	-.137
Engagement										
Part-time job	134	52.90	8.79	229	39.83	14.86	71	-1.276	.214	-.233
Smoking ban	168	70.65	11.34	233	43.11	16.68	41	-2.596	.008	-.474
Graduation										
Part-time job	139	46.25	8.83	174	32.58	9.20	39	-2.684	.006	-.490
Smoking ban	129	52.97	9.06	164	52.85	11.61	97	-.132	.914	-.024

Table 2: *Appraisal resources employed by L1 and L2 writers*

We now explore each of the main *appraisal* categories (*attitude, engagement* and *graduation*) in turn.

7.2. Attitude

Table 3 shows that for the part-time job task, *appreciation* was the predominant type of *attitude* as the normalised median frequency of *appreciation* is higher than the other two types namely *affect* and *judgement*. This suggests both groups preferred evaluating phenomena to expressing their feelings or judging behaviour in evaluation. Both L1 and L2 writers frequently used epithets to evaluate phenomena when expressing their evaluative stance:

(1) A (+*appreciation*) good part-time job enables students to work as a team. [HKL2E]

Significant L1/L2 differences were sourced in the use of *authorial evaluation*, with the NES using this function infrequently (see Excerpt 2), while the HKL2E writers never did so, using *non-authorial evaluation* in a small number of cases instead.

(2) Finally, I just decided to go ahead and do it and I am (+*affect, authorial evaluation*) glad I did. [NES]

In terms of explicitness of *attitude*, both groups preferred inscribed evaluation, although HKL2E writers tended to state their evaluative stance on this topic with a positive *attitude* (raw=182, normed=34.69) more than NES writers (raw=65, normed=30.04).

(3) It [A part-time job] is a gateway to them to be a (+*attitude*) successful person. [HKL2E]

	L1			L2			Mann–Whitney *U* test			
	f	*f* per 1,000 words	median absolute deviation	*f*	*f* per 1,000 words	median absolute deviation	U	*z*	*p*	Pearson's *r*
Affect										
Authorial evaluation	11	0	0	0	0	0	60	-2.97	.003	-.540
Non-authorial evaluation	2	0	0	7	0	0	101	.063	.950	-.011
Judgement	28	30.84	9.88	58	9.22	6.85	98	-.089	.929	-.091
Appreciation	73	52.90	8.79	229	39.83	14.86	71	-1.276	.214	-.233
Explicitness										
Inscribed	98	38.02	6.03	223	41.50	8.46	123	1.01	.312	-.185
Invoked	16	3.80	3.80	228	1.69	1.69	97	-.141	.888	-.026
Polarity										
Positive	65	30.04	3.55	182	34.69	7.63	142	1.84	.065	-.337
Negative	49	16.13	7.55	69	12.08	4.51	88	-.528	.598	-.096

Table 3: *Attitude resources employed by L1 and L2 writers ('part-time job' task) (alpha=0.00625)*

When discussing the smoking ban, *appreciation* is still the preferred means of conveying *attitude* as the normalised median frequency is the highest compared to the other two types. The use of *affect* and *judgement* resources is much lower than those seen for the 'part-time job' task in both L1 and L2 essays. Both groups therefore made less personal expressions of *attitude* and focused more on facts, for example the pros and cons of having a smoking ban, in their arguments.

(4) […] smoking is terribly (-*appreciation*) unhealthy. [NES]

(5) Smokers in the restaurant will provide second-hand smoke and make the environment (-*appreciation*) unpleasant. [HKL2E]

Both groups still rarely used *authorial evaluation* when dealing with this topic, but HKL2E writers made a few attempts.

(6) Smoking section contains smokers and tobaccos ashes which will (-*affect, authorial evaluation*) discomfort me. [HKL2E]

In particular, we see significant differences between L1/L2 writers in the polarity of their attitudes in the discussion of a smoking ban, with L1 writers using significantly more instances of positive *attitude*, and with L2 writers using significantly more instances of negative *attitude*. It was also found that L1 writers deployed attitudinal resources, both positive and negative *attitude*, to express their opinions not directly related to smoking bans or smoking.

(7) I have never been to Japan, but I would definitely like to travel there one day, and nothing ruins the experience of a (+*attitude*) good meal with your friends or family at a restaurant like the smell of cigarette smoke. [NES]

Excerpt 7 shows that L1 writers more frequently built up their arguments in different ways, unlike L2 writers who generally preferred pointing out the disadvantages of smoking through supporting evidence, as illustrated in Excerpt 8.

(8) When people smoke, ashes may drop onto our food being prepared and is very (-*attitude*) harmful to our health. [HKL2E]

	L1			L2			Mann–Whitney *U* test			
	f	*f* per 1,000 words	median absolute deviation	*f*	*f* per 1,000 words	median absolute deviation	U	z	p	Pearson's *r*
Affect										
Authorial evaluation	7	0	0	4	0	0	76	-1.35	.175	-.248
Non-authorial evaluation	2	0	0	7	0	0	90	.565	.572	-.103
Judgement	5	0	0	15	0	0	92	-.389	.697	-.071
Appreciation	74	29.41	7.15	170	31.65	9.09	76	-1.05	.291	-.193
Explicitness										
Inscribed	78	31.75	4.35	185	41.25	10.96	70	1.32	.187	.241
Invoked	10	0	0	11	0	0	83	-.922	.356	-.168
Polarity										
Positive	44	18.12	3.74	56	9.48	3.85	44	-2.44	.015	.446
Negative	44	18.02	8.31	140	27.72	11.41	49	-2.22	.026	.406

Table 4: *Attitude resources employed by L1 and L2 writers ('smoking ban' task) (alpha=0.00625)*

7.3. Engagement

	L1			L2			Mann–Whitney *U* test			
	f	*f* per 1,000 words	median absolute deviation	*f*	*f* per 1,000 words	median absolute deviation	U	*z*	*p*	Pearson's *r*
Monoglossic	3	0	0	9	0	0	98	-.136	.892	-.025
Heteroglossic	131	49.08	5.31	220	39.83	14.04				
Contract										
Disclaim										
Deny	36	13.30	7.68	54	8.05	4.66	71	1.27	.202	-.233
Counter	22	7.39	3.85	34	7.35	2.89	80	-.861	.389	-.157
Proclaim										
Concur										
Affirm	2	0	0	5	0	0	99	-.063	.950	-.011
Concede	1	0	0	0	0	0	90	-1.41	.157	-.258
Pronounce	8	3.61	0.80	14	0	0	86	-.062	.531	-.114
Justify	16	4.46	1.35	38	3.93	3.93	84	-.704	.481	-.130
Expand										
Entertain	39	16.19	3.15	59	8.15	8.15	69	-1.36	.172	-.250
Attribute										
Acknowledge	7	1.85	1.85	11	0	0	83	-.866	.387	-.158
Distance	0	0	0	5	0	0	80	-1.48	.137	-.272

Table 5: *Engagement resources employed by L1 and L2 writers ('part-time job' task) (alpha=0.005)*

Table 5 shows that for the 'part-time job' task, there were no significant differences reported between L1/L2 groups in the use of *engagement* resources, although some differences are apparent from the effect size scores. Overall, NES writers employed a non-significantly higher frequency of *engagement* resources than HKL2E writers for this task, with NES writers occasionally producing overt authorial stance through *pronounce* devices (raw=8, normed=3.61) such as 'I agree' and 'the fact', as well as clausal intensifiers such as 'indeed' and 'really' to indicate their position:

(9) (*pronounce*) The fact of the matter is that we are already busy enough learning. [NES]

Although the frequency of instances of *proclaim* was similar in L1 essays and L2 essays, it was found that HKL2E writers used *justify* as the main *proclaim* strategy (raw=38, normed=3.93) to *justify* a proposition by giving reasons to seek to limit the set of options for responses by others. This rhetorical strategy, shown in Excerpt 10, was straightforward and was achieved by using simple connectives such as 'because', 'as' and 'since', etc.

(10) However, it is difficult for university students to have a full-time job to earn money (*justify*) because they have to study. [HKL2E]

	L1			L2			Mann–Whitney *U* test			
	f	*f* per 1,000 words	median absolute deviation	*f*	*f* per 1,000 words	median absolute deviation	U	*z*	*p*	Pearson's *r*
Monoglossic	0	0	0	0	0	0				
Heteroglossic	188	70.65	11.34	235	43.11	16.68				
Contract										
Disclaim										
Deny	43	16.88	4.17	56	8.57	4.49	44	-2.46	.014	-.450
Counter	33	16.09	4.39	50	9.86	2.36	71	-1.60	.107	-.289
Proclaim										
Concur										
Affirm	1	0	0	2	0	0	95	-.048	.647	-.084
Concede	10	0	0	2	0	0	67	-2.07	.038	-.379
Pronounce	11	4.31	0.25	16	3.46	3.4	65	-1.56	.117	-.286
Justify	11	0	0	34	8.01	4.55	71	-1.26	.207	-.230
Expand										
Entertain	78	21.86	12.20	53	5.36	4.95	41	-2.56	.009	-.474
Attribute										
Acknowledge	7	4.17	2.33	20	3.66	3.66	97	-.138	.895	-.025
Distance	0	0	0	2	0	0	90	1.01	.309	-.186

Table 6: *Engagement resources employed by L1 and L2 writers ('smoking ban' task) (alpha=0.0055)*

As mentioned, significant differences were found in the use of *engagement* resources between L1 and L2 writers for the 'smoking ban' task (Table 6). HKL2E writers used significantly fewer *heteroglossic* resources, in particular dialogic expansion, than L1 English writers. We also note that L1 writers are significantly more likely than L2 writers to *disclaim* the topic at hand through denial in dialogic discussion:

(11) I have been to a few restaurants where the non-smoking area really (*deny*) didn't seem that much different from the smoking area … [NES]

The higher frequency of *entertain* in dialogic expansion in L1 texts shows that NES writers opened up space for negotiation of other voices and perspectives around a negative issue between the writer and readers:

(12) Government (*entertain*) may be surprised at the amount of support that they will receive for introducing the bans. [NES]

It was also observed that NES writers limited the dialogic space utilising *concede* devices more frequently than their counterparts to develop their arguments:

(13) (*concede*) Sure, we eat a lot of fried foods and drink a lot of sweet tea… [NES]

Although not statistically significantly different, the only resource that HKL2E writers made greater use of than NES writers was *justify* in dialogic contraction (raw=34, normed=8.01). L2 writers appeared to use overt reasoning as the main strategy to persuade readers to support their arguments (example 14).

(14) (*justify*) *As we cannot avoid smoke from passing to non-smoking area, it is not a solution to passive smoking.* [HKL2E]

7.4. Graduation

	L1			L2			Mann–Whitney *U* test			
	f	*f* per 1,000 words	median absolute deviation	*f*	*f* per 1,000 words	median absolute deviation	U	*z*	*p*	Pearson's *r*
Force										
Upscale	106	41.89	5.32	144	28.12	8.82	47	-2.33	.020	-.426
Downscale	12	3.98	3.31	24	4.26	4.26	98	0.06	.947	-.012
Focus										
Soften	3	0	0	3	0	0	86	-.594	.553	-.155
Sharpen	18	1.65	1.65	3	0	0	59	-2.28	.022	-.418

Table 7: *Graduation resources employed by L1 and L2 writers ('part-time job' task) (alpha=0.0125)*

As mentioned, NES writers used significantly more *graduation* resources than HKL2E writers for amplification (Table 7), as shown in the dominant use of sharpen resources when producing *focus*, using significantly more semantic resources for intensification through *upscale*. Excerpt 15 shows how NES writers used *focus* to soften the non-gradable experiential meaning, a piece of succulent steak.

(15) It is (-*focus*) somewhat like a piece of succulent steak, studying is like its protein. [NES]

The higher number of instances of *upscale* found in L1 texts suggests that NES writers were more likely to intensify their claims in argumentation than HKL2E writers (Excerpt 16):

(16) There are (+*force*) plenty of ways to find a part-time job... [NES]

In contrast, HKL2E writers relied mostly on the use of attitudinal resources to amplify meanings and indicate their positioning. They made very few attempts to use *focus* as a strategy to sharpen categorical meanings compared to their counterparts.

While the overall frequencies of *graduation* devices between L1 and L2 groups was similar for the 'smoking ban' task (Table 8), significant differences are still reported for the use of softening *focus* devices (although this is likely because there are no instances in the L2 data). Although the differences for other categories are not statistically significant, HKL2E writers made greater use of *downscale force* on this topic (raw=37, normed=4.37) compared with NES writers. HKL2E writers employed *downscale* as a strategy to be less assertive about their claims to seek alignment with readers because they had to rely on their knowledge about the topic when formulating their arguments, as in Excerpt 17:

(17) By having (-*force*) a little private space, it can make a huge difference to our dining experience. [HKL2E]

	L1			L2			Mann–Whitney *U* test			
	f	*f* per 1,000 words	median absolute deviation	*f*	*f* per 1,000 words	median absolute deviation	U	*z*	*p*	Pearson's *r*
Force										
Upscale	119	46.82	13.26	216	42.49	8.41	91	-.396	.692	-.072
Downscale	8	0	0	37	4.37	4.37	65	-1.59	.112	-.291
Focus										
Soften	10	1.43	1.43	0	0	0	80	2.03	.042	-.371
Sharpen	0	0	0	0	0	0		n/a		

Table 8: *Graduation resources employed by L1 and L2 writers ('smoking ban' task) (alpha=0.0167)*

8. Discussion

This study involved a CIA analysis of *appraisal* resources in L1 and L2 English argumentative essays, to determine variation on how L1 and L2 writers express *appraisal* in argumentative essays produced under the same conditions and on the same topic. The study represents a rare (albeit relatively small) corpus-informed instantiation of the *appraisal* framework, increasing the representativeness of the findings for L1 and L2 writers.

The findings of the study, supported by the non-parametric tests applied, suggest that NES writers and HKL2E writers of an intermediate English proficiency level specific to this dataset deployed all three domains of interpersonal resources within the *appraisal* model to construct their evaluative stance in these argumentative essays. Resources of *engagement* were used most frequently, followed by *graduation* and *attitude*. The genre, one-sided argumentative essays, presumably impacted the distribution of evaluative resources given that the quality and persuasiveness of arguments are crucial in this genre. This accounts for the frequent use of *engagement* resources, since the writers developed their arguments by introducing their authorial voice and engaging with other voices. *Graduation* resources were also often deployed to vary the strength of authors' claims to increase the persuasiveness of their arguments. The use of *attitude* resources contributed the least to the construction of evaluative stance, suggesting that NES writers and HKL2E writers at this proficiency level in our dataset are aware that effective authorial stance cannot be construed only through expression of feelings, assessment of people's behaviour, or evaluation of entities, in this genre.

Some similar characteristics of the use of evaluative resources across the two writing tasks were detected. First, the writers preferred expressing their attitude directly. In one-sided argumentative writing, the authors explicitly stated their stance towards a proposition, leading to a higher frequency of explicit evaluation. Second, authorial intrusion was minimal, as reflected in the relative underuse of *authorial evaluation* in all texts. Although the writers made explicit attitudinal choices, they ascribed their arguments to broader community views instead of their personal feelings to remain objective in the discussion. Third, the writers used a limited range of dialogic contraction and dialogic expansion resources. Negation, introduction of counterarguments and opening up space for a range of possible voices were the common persuasive strategies identified in all essays. These strategies were easily realised via individual lexical items (e.g., 'no', 'nothing'), connectives (e.g., 'however', 'but') and common expressions (e.g., 'in my view') respectively. The writers seldom highly endorsed a proposition probably because they tried to minimise subjectivity by allowing alternative voices. Lastly, the writers relied heavily on resources of *force* to vary the strength of arguments through intensification and quantification, regardless of topics in argumentation.

It was also found that, on the whole, evaluative stance was realised by a more 'even' combination of the three main domains of the *appraisal* system in L2 essays, while *engagement* featured more heavily in L1 essays. HKL2E writers deployed *attitude* resources (particularly *appreciation*) the most, with NES writers using them the least. This suggests that the construction of evaluative stance in these L2 essays is largely framed by the authors' feelings, including emotional reactions, judgement of behaviour and evaluation of entities. In contrast, NES writers showed a stronger preference for employing various means to negotiate relationships of alignment and disalignment with readers via *engagement* devices. They also deployed more *graduation* resources than HKL2E writers to vary the strength of evaluation. The different orientations demonstrate that NES writers often varied the intensity of their arguments positioned in a text and aligned readers into that value position (Martin and White, 2005), while HKL2E writers focused their evaluation on value of things such as the pros or the cons of having a part-time job or smoking. To summarise, NES writers used different dialogic resources to engage with the topic and their readers to increase persuasiveness of their textual voice, while HKL2E writers put forward their claims directly with less interaction and engagement with readers.

Specifically in terms of *attitude*, both NES and HKL2E writers indicated a strong preference for evaluating entities over expressing their feelings or judging people's behaviour or character, as shown in the prevalent use of *appreciation* in *attitude*. Both sets of writers maintain objectivity by making less personal expression and focusing more on facts in their argumentation. This finding echoes the results of previous research (e.g., Hood, 2004, 2005, 2006; Derewianka, 2007) that *appreciation* is the most commonly used form of *attitude* in writing. In Hood's (2005) study, *appreciation* was predominant in student texts in the introductory sections of academic research papers. One possible explanation of the dominant use of *appreciation* in the present study is that writers who are of an intermediate English proficiency level have passed the stage where the use of *affect* is predominant in evaluation (Derewianka, 2007). Derewianka finds that there is no recourse to *affect* and no overt *judgement* in student texts in low-proficiency tertiary writing, and those more proficient writers rely more heavily on *appreciation* than *affect* or *judgement* to construct their evaluative stance. However, NES writers attempted to use *affect* resources, which were rarely used by the HKL2E writers in our study, to construct a more personal response to the writing prompts. Our L2 finding here contradicts Hood's (2005) finding that L2 writers construct a more personalised expression of evaluation through the use of *affect*. Due to a greater emphasis on objectivity in many L2 pedagogical argumentative writing materials (Swain, 2007), HKL2E writers might avoid expressing personal feelings to remain objective in argumentation. Other potential reasons for the underuse of *affect* by HKL2E writers is because of the disproportionate attention on *appreciation* such writers have received in L2 pedagogical writing materials (Hyland,

1994) which emphasise the importance of objectivity in persuasive writing (Swain, 2007), or that the HKL2E writers at this proficiency might already have access to a range of attitudinal resources in their interlanguage. For instance, epithets were often used when discussing the advantages or disadvantages of having a part-time job for college students. However, this is because L2 writers might find it easier to use a single lexical item, for example an adjective, in evaluation as the syntactic structure is less complicated. Another reason for the frequent use of *appreciation* may be due to L1 transfer. The use of epithets is typical in argumentative essays written in Standard Cantonese, their first language. Consequently, HKL2E writers might adopt a similar strategy in their L2 production. Excerpt 18 illustrates the use of epithets in Cantonese, which resembles a sentence structure extracted from a L2 essay about part-time jobs.

(18) Choosing (+*appreciation*) appropriate types of part-time job is (+*appreciation*) important too. [HKL2E]

選擇	合適的	兼職	類別	同樣	重要
syun2 zaak6	hap6 sik1 dik1	gim1 zik1	leoi6 bit6	tung4 joeng6	zung6 jiu3
choose	appropriate	part-time job	type	is	important

Regarding the types of *affect*, although the total number of instances of *authorial evaluation* in all texts remained low, there were slightly more instances identified in L1 than L2 essays. The absence of persuasive supporting evidence might account for the use of *authorial evaluation* among NES writers in this ICNALE dataset. They did not have access to other sources or reference materials when they produced the essays, so they occasionally expressed their feelings towards the topic ('part-time job'), which was more relevant to their real-life situation when construing their authorial voice. As suggested by Coffin and Hewings (2004), an excess of authorial intrusion may weaken rather than strengthen the overall argument. HKL2E writers might also deliberately avoid using this device to minimize authorial intrusion. On the other hand, Hyland (1994) observes that L2 pedagogical writing materials give disproportionate attention to different aspects of evaluation other than *authorial evaluation*. L1 transfer might also account for the extensive use of positive and negative adjectives, which are commonly used in argumentative essays written in Chinese. The HKL2E writers pointed out the disadvantages of smoking via such adjectives as supporting examples to justify their proposition, which led to a significantly higher frequency of negative *attitude* in L2 than L1 texts. Excerpt 19 is a translation of English into Chinese to demonstrate how HKL2E writers would express the same idea in their L1.

(19) Smoking causes a wide variety of (-*attitude*) adverse effects to human health. [HKL2E]

吸煙	對	人體	健康	產生	很多
kap1 jin1	deoi3	jan4 tai2	gin6 hong1	caan2 saang1	han2 do1
smoking	to	human	health	cause	many

不同的	負面	影響
bat1 tung4 dik1	fu6 min6	jing2 hoeng2
a wide variety of	adverse	effects

In terms of *engagement*, NES writers used a broader range of linguistic resources to position their values and construct alternative viewpoints in a text (Martin and White, 2005), evidenced in the significantly higher number of instances of *engagement* found in L1 than L2 essays, as well as more varied patterns of dialogic expansion and dialogic contraction resources in *heterogloss*. HKL2E writers were more direct when making claims depending on a relatively limited range of *heteroglossic* resources. This finding is similar to Coffin and Hewings' (2004) observation that Chinese writers are more direct in putting forward claims in the L2 because of their limited repertoire of interpersonal resources of their interlanguage. For L1 texts, *deny* was the predominant dialogic contraction resource in *disclaim*. According to Martin and White (2005), negation (*deny*) is a resource for invoking an alternative viewpoint so as to reject it. This directed outwards and away from the current writer-reader relationship since the writer indicated a disalignment with some third party. Nevertheless, Martin and White (2005) claim that denial is corrective rather than confrontational, and will enhance solidarity between the writer and readers if readers do not reject the particular viewpoint being advanced. In addition, Coffin and Hewings (2004) claim that acknowledging diversity is effective because it accumulates additional authority for the writer's views and opens them up to alternative positions and contestation, increasing persuasiveness. The additional finding that *entertain* rather than *attribute* in dialogic expansion was commonly deployed by both groups to open up space for alternative viewpoints is interesting in that *entertain* was used by HKL2E writers less frequently. This finding suggests that their claims were more assertive in general. The result corresponds to Coffin and Hewings' (2004) finding that it is common for novice writers to use a limited range of interpersonal resources such as strongly affirmed opinions.

In terms of *graduation*, both L1 and L2 groups relied heavily on *force* to amplify experiential meanings. Although NES writers used significantly more instances of *focus* to sharpen non-gradable attitudinal meanings, the overall number identified in all texts was much lower than *force*. One possible explanation for its underuse by the HKL2E group is that sharpening and softening values to maximize or minimize the impact of authorial voice might be beyond their reach in their language proficiency. This seems to cohere with Derewianka's (2007) observation that L2 students use pre-modifiers and resources of quantification in *force* more frequently when they are less proficient and they are only able to use *force* and *focus* strategically towards an early tertiary stage. When discussing part-time work, NES writers increased the intensity of their arguments by multiple instances of *graduation*, particularly *upscale* in *force*. They deployed a wider range of *graduation* resources to encode their subjective positioning, and let readers interpret the

position being advanced in the text (Hood, 2006) as *force* interacts with *attitude* and have associated effects with respect to alignment and solidarity (Martin and White, 2005). HKL2E writers, on the other hand, used mostly attitudinal resources via *appreciation* to amplify experiential meanings and indicate their positioning. Moreover, NES writers used significantly more instances of *focus*, especially sharpen, than their counterparts. Sharpening values can maximise the effect of the authorial voice in the value positioned in a text. HKL2E writers should therefore learn to manage the full range of *graduation* resources recommended by Hood (2006) to position their evaluative stance more effectively.

9. Conclusion

The functional perspective of this paper reveals comparative and divergent patterns of evaluative choices made by L1 and L2 writers. We suggest that the relative under/over use of certain features across the three *appraisal* domains by the L2 writers found in this study should help to inform writing teachers about the required repertoire of interpersonal meaning-making resources that students need to acquire to extend their repertoire. Explicit writing instruction on these features is recommended to raise awareness of L1/L2 divergence, and a principle of progression should be adopted in curriculum, syllabus and teaching material design when introducing interpersonal resources to students. Lastly, teachers should provide specific feedback on the evaluative aspect of language use in student writing.

One of the major limitations of the study is its small sample size extracted from a learner corpus. *Appraisal* is a comprehensive analytical framework of semantic resources of interpersonal meaning. Every instance of implicit and explicit evaluation are highlighted and analysed. Such a word-to-word analysis approach poses constraints to the amount of data to be analysed manually, so the scale of the study is necessarily small. ICNALE texts are also quite short, and it is possible that the writers have not been able to demonstrate full command of *appraisal* resources across such short texts. The research findings should therefore be interpreted against the backdrop of the language backgrounds of our dataset. We feel that even with this small sample, the non-parametric tests have still shown significant differences across NES and HKL2E groups that should only be further intensified (rather than negated) with an increased sample. Besides a full *appraisal* analysis, another possible research area is the impacts of genres on evaluation. Future studies could involve other types of persuasive writing, such as speeches and editorials, to explore the construction of evaluative stance from multiple perspectives. Despite the limitations of the study, the study demonstrated how corpora could be used to look at evaluative language, and that as corpus linguistics and SFL can be more closely aligned (Flowerdew, 2004), future research under an SFL paradigm can benefit from corpus analysis.

References

Bakhtin, M. M. (1986). *Speech Genres and Other Late Essays.* Austin, TX: University of Texas Press.

Biber, D. & Finegan, E. (1988). Adverbial stance types in English. *Discourse Processes, 11*(1), 1–34. doi:10.1080/01638538809544689

Bley-Vroman, R. (1983). The comparative fallacy in interlanguage studies: the case of systematicity. *Language Learning, 33*(1), 1–17. doi:10.1111/j.1467-1770.1983.tb00983.x

Bruce, I. (2014). Expressing criticality in the literature review in research article introductions in applied linguistics and psychology. *English for Specific Purposes, 36,* 85–96. doi:10.1016/j.esp.2014.06.004

Chan, A. Y. W. (2004). Syntactic transfer: evidence from the interlanguage of Hong Kong Chinese ESL learners. *The Modern Language Journal, 88*(1), 56–74. doi:10.1111/j.0026-7902.2004.00218.x

Cicchetti, D. V. (1994). Guidelines, criteria, and rules of thumb for evaluating normed and standardized assessment instruments in psychology. *Psychological Assessment, 6*(4), 284–290. doi:10.1037/1040-3590.6.4.284

Coffin, C., & Hewings, A. (2004). IELTS as preparation for tertiary writing: distinctive interpersonal and textual strategies. In L. Ravelli & R. Ellis (Eds.), *Analysing Academic Writing: Contextualised Frameworks* (pp. 153–171). London: Continuum.

Council of Europe. (2001). *A Common European Framework of Reference for Learning, Teaching and Assessment.* Cambridge: Cambridge University Press.

Crosthwaite, P. (2016). L2 English article use by L1 speakers of article-less languages: A learner corpus study. *International Journal of Learner Corpus Research, 2*(1), 68 –100, doi:10.1075/ijlcr.2.1.03cro

Crosthwaite, P., Cheung, L. M. L. & Jiang, F. K. (2017). Writing with attitude: stance expression in learner and professional dentistry research reports. *English for Specific Purposes, 46,* 107–123, doi:10.1016/j.esp.2017.02.001

Derewianka, B. (2007). Using *appraisal* theory to track interpersonal development in adolescent academic writing. In A. McCabe, M. O'Donnell & R. Whittaker (Eds.), *Advances in Language and Education* (pp. 142–165). London: Continuum.

Dörnyei, Z. (2007). *Research Methods in Applied Linguistics: Quantitative, Qualitative, and Mixed Methodologies.* Oxford: Oxford University Press.

Flowerdew, L. (1998). Corpus linguistic techniques applied to textlinguistics. *System, 26*(4), 541–552. doi:10.1016/S0346-251X(98)00039-6

Flowerdew, L. (2004). The problem-solution pattern in apprentice vs. professional technical writing: an application of appraisal theory. In G. Aston, S. Bernardini, & D. Stewart (Eds.), *Corpora and Language Learners* (pp. 125–135). Amsterdam: John Benjamins.

Geng, Y., & Wharton, S. (2016). Evaluative language in discussion sections of doctoral theses: similarities and differences between L1 Chinese and L1 English writers. *Journal of English for Academic Purposes, 22*, 80–91. doi:10.1016/j.jeap.2016.01.001

Gilquin, G. (2001). The Integrated Contrastive Model: spicing up your data. *Languages in Contrast, 3*(1), 95–213.

Granger, S. (1996). From CA to CIA and back: an integrated approach to computerized bilingual and learner corpora. In K. Aijmer, B. Altenberg, & M. Johansson. (Eds.), *Languages in Contrast. Papers from a Symposium on Text-Based Cross-Linguistic Studies* (pp. 37–51). Lund: Lund University Press.

Granger, S. (2002). A bird's-eye view of learner corpus research. In S. Granger, J. Hung, & S. Petch-Tyson (Eds.), *Computer Learner Corpora, Second Language Acquisition and Foreign Language Teaching* (pp. 3–33). Amsterdam: John Benjamins.

Granger, S. (2004). Computer learner corpus research: current status and future prospects. In U. Connor & T. A. Upton (Eds.), *Applied Corpus Linguistics: A Multidimensional Perspective* (pp. 123–145). Amsterdam: Rodopi.

Granger, S. (2015). Contrastive interlanguage analysis: a reappraisal. *International Journal of Learner Corpus Research, 1*(1), 7–24. doi:10.1075/ijlcr.1.1.01gra

Green, C. F., Christopher, E. R. & Lam, J. K. M. (2000). The incidence and effects on coherence of marked themes in interlanguage texts: a corpus-based enquiry. *English for Specific Purposes, 19*(4), 99–113. doi:10.1016/S0889-4906(00)00015-6

Halliday, M. A. K. (1994). *An Introduction to Functional Grammar* (2nd Ed.). New York: Routledge.

Hong, H. & Cao, F. (2014). Interactional metadiscourse in young EFL learner writing: A corpus-based study. *International Journal of Corpus Linguistics, 19*(2), 201–24. doi:10.1075/ijcl.19.2.03hon

Hood, S. (2004). *Appraising Research: Taking a Stance in Academic Writing* (Unpublished doctoral dissertation). University of Technology, Sydney.

Hood, S. (2005). Managing attitude in undergraduate academic writing: a focus on the introductions to research reports. In L. Ravelli & R. Ellis (Eds.), *Analysing Academic Writing: Contextualized Frameworks* (pp. 24–44). London: Continuum.

Hood, S. (2006). The persuasive power of prosodies: radiating values in academic writing. *Journal of English for Academic Purposes, 5*(1), 37–49. doi:10.1016/j.jeap.2005.11.001

Hood, S. (2010). *Appraising Research: Evaluation in Academic Writing*. London: Palgrave Macmillan.

Hunston, S. (2000). Evaluation and the planes of discourse: status and value in persuasive texts. In S. Hunston & G. Thompson (Eds.). *Evaluation in Text: Authorial Stance and the Construction of Discourse* (pp. 176–207). Oxford: Oxford University Press.

Hunston, S. & Thompson G. (2000). Evaluation: an introduction. In S. Hunston & G. Thompson (Eds.). *Evaluation in Text: Authorial Stance and the Construction of Discourse* (pp. 1–27). Oxford: Oxford University Press.

Hyland, K. (1994). Hedging in academic writing and EAP textbooks. *English for Specific Purposes, 13*(3), 239–256. doi:10.1016/0889-4906(94)90004-3

Hyland, K. & Milton, J. (1997). Qualification and certainty in L1 and L2 students' writing. *Journal of Second Language Writing, 6*(2), 183–205. doi:10.1016/S1060-3743(97)90033-3

Hyland, K. & Tse, P. (2004). Metadiscourse in academic writing: a reappraisal. *Applied Linguistics, 25*(2), 156–177. doi:10.1093/applin/25.2.156

Hyland, K. (2005). Stance and engagement: a model of interaction in academic discourse. *Discourse Studies, 7*(2), 173–192. doi:10.1177/1461445605050365

Ishikawa, S. (2011). A new horizon in learner corpus studies: the aim of the ICNALE project. In G. Weir, S. Ishikawa, & K. Poonpon (Eds.), *Corpora and Language Technologies in Teaching, Learning and Research* (pp. 3–11). Glasgow: University of Strathclyde Press.

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* (pp. 91–118). Kobe: Kobe University School of Languages and Communication.

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

Neff-van Aertselaer, J. (2016). Learner corpora and discourse. In S. Granger, G. Gilquin & F. Meunier (Eds.) *The Cambridge Handbook of Learner Corpus Research* (pp. 255–279). Cambridge: Cambridge University Press.

O'Donnell, M. (2008). The UAM CorpusTool: software for corpus annotation and exploration. In C. M. Bretones Callejas, J. F. Fernández Sánchez, J. R. Ibáñez Ibáñez, M. E. García Sánchez, M. E. Cortés de los Ríos, S. Salaberri Ramiro *et al.* (Eds.). *Applied Linguistics Now: Understanding Language and Mind / La Lingüística Aplicada Hoy: Comprendiendo el Lenguaje y la Mente* (pp. 1433–1447). Almería: Universidad de Almería.

Paltridge, B. (1997). *Genre, Frames and Writing in Research Settings.* Amsterdam: Benjamins.

Sinclair, J. (2004). *Trust the Text: Language, Corpus and Discourse*. London: Routledge.

Swain, E. (2007). Constructing an effective 'voice' in academic discussion writing: an appraisal theory perspective. In A. McCabe, M. O'Donnell, & R. Whittaker (Eds.), *Advances in Language and Education* (pp. 166–184). London: Continuum.

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

Thompson, G. & Zhou, J. (2000). Evaluation and organization in text: the structuring role of evaluative disjuncts. In S. Hunston & G. Thompson (Eds.). *Evaluation in Text: Authorial Stance and the Construction of Discourse* (pp. 121–141). Oxford: Oxford University Press.

White, P. R. R. (2003). Beyond modality and hedging: a dialogic view of the language of intersubjective stance. *Text, 23*(2), 259–284. doi:10.1515/text.2003.011

Ye, S. (2016). How online news media portray Vladimir Putin differently during the Ukrainian crisis: comparative study of America and Chinese online news media. Unpublished MA dissertation, University of Hong Kong *Theses Online (HKUTO)*. doi:10.5353/th_b5812784