Using Assessment Data to Investigate the Relationships Between Speaking and Writing

Amy Kwai Mun Cheung
Hong Kong Examinations and Assessment Authority
Flora Fung Yin Leung
Education Bureau, Hong Kong

Overview

This paper discusses the relationships between speaking and writing in the study: 'Using Assessment Data (Including TSA Data) to Enhance the Learning and Teaching of Speaking and Writing (English Language) at Key Stage Three' , a study jointly conducted by the Education Bureau (EDB) and Hong Kong Examinations and Assessment Authority (HKEAA) in 2011.

Based on the findings, an online corpus-related self-learning resource with a diagnostic tool has been developed to help identify and analyse students' difficulties in English learning and help them improve.

Literature Review

In the field of language proficiency testing, trans-subjective[1] measures of countable features can be employed as a check on the validity of ratings of assessment criteria, e.g., Banerjee, et al., (2007) and Hawkey and Barker (2004) use 'syntactic complexity' and 'vocabulary richness' for trans-subjective measurement of 'sophistication of language', 'grammatical accuracy' for 'language', and 'cohesive devices' for 'organization and cohesion'. Measures using countable features are referred to as verifiable quantitative measures (VQM) (Cheung, 2010). The measures include the length of T-units and the number of clauses per T-unit, a measure of syntactic complexity. Syntactic complexity has been investigated in L2 writing studies as well as in L2 speech data (Crookes, 1989; Ortega, 1999; Skehan & Foster, 1999).

[1] Trans-subjective is used to describe information which once would have been termed 'objective'. However, as Foucault (1974) points out that there is no such thing as objectivity, the closest we can get to objective measures are trans-subjective measures.

Florez and Hadaway (1987) state that oral language development can have an effect on writing behaviour, but that oral language proficiency scores may not indicate what to expect from students in written composition. Kim (2000) points out that the stages in L1 and L2 acquisition might not be the same. Hubert (2008) investigates the relationship between writing and speaking in the U.S. university Spanish language classroom. He finds weak correlation between speaking and writing at beginning levels of study, and much stronger correlation at the intermediate and advanced levels. Zhu (2007) conducts a study among 40 randomly selected college-level ESL (English as a second language) students in one American university. It is found that the college-level students with good speaking skills have good writing skills and vice versa (r values 0.67-0.86).

Purpose

This study involved a stratified sample of 180 Secondary 3 (S3) (Grade 9) student performances (spoken and written performances of the same students) from Hong Kong. Its aims were to investigate: 1) rated and counted features of spoken and written performances in English Language testing; 2) students' performances in speaking and writing; 3) students' perceptions on their spoken and written performances

Research Questions (RQ)

RQ1. To what extent do the sub-constructs within a skill and across two skills correlate?

RQ2. To what extent do the rated and counted features of the sub-constructs within a skill and across two skills correlate?

RQ3. What can we deduce from comparing student performances in writing and speaking tasks?

Participants

The Project Team consisted of two investigators, Senior Curriculum Development Officer of the Assessment & HKEAA Section of EDB and Manager of Education Assessment Services Division of HKEAA.

One hundred and eighty[2] Secondary 3 (Grade 9) students were selected from six schools on a stratified random base taking speaking and writing assessments. Each student took a writing task and an individual presentation task with matching topics. The Project Team conducted and recorded the assessment of individual presentation.

The participating school teachers invigilated the writing assessment within one week of the oral assessment. Three experienced teachers were trained and appointed as raters to rate both writing and speaking tasks, each rating 120 writing scripts and 120 oral tasks.

Methods of Analysis

Having collected a stratified sample of 180 Secondary 3 student performances (spoken and written), students' performances were rated by three experienced teacher raters. 'Fair average' scores from raters' ratings were then obtained using *FACETS* (Linacre, 1991-2008). With the spoken performances transcribed, the Team selected, analyzed and annotated three sets of students' spoken performances and written performances of the same students.

The Project Team calculated counted features of student performances either by humans or by computer software, i.e. verifiable quantitative measures (VQM) (Cheung, 2010). Human counted VQM included 'T-Unit index' (syntactic complexity), 'grammatical accuracy index', 'spelling', 'punctuation' and 'capitalization' for 'language' , 'pronunciation accuracy index' for 'pronunciation' and 'fluency index' for 'delivery'. Computer software included 'types' 'tokens', 'type -token ratio' using *RANGE* (Heatley, et al., 2002) and 'lexical diversity (D)' using *vocd* (Malvern, et al., 2004) for the sub-construct 'vocabulary'. Pearson's 'r' was used to calculate correlations among sub-constructs within and across two skills. To rid a correlation coefficient from the weakening effect of measurement error, the following formula was used for the correlations between rated features across two skills:

$$r_{x'y'} = \frac{r_{xy}}{\sqrt{r_{xx}r_{yy}}}.$$

[2] A total of 173 S3 student performances in speaking and writing were collected since a total of seven students were absent on the day of written assessment.

Findings and Discussions

RQ1. To what extent do the sub-constructs within a skill and across two skills correlate?

1.1 Correlation of overall ratings and between ratings of sub-constructs within a skill

The results in Tables 1.1 and 1.2 show high to very high correlation (0.75-0.96) within the sub-components of the two macro skills in question, i.e. speaking and writing.

In individual presentations, the correlation between 'vocabulary and language patterns' (VL) and 'pronunciation and delivery' (PD) was very high ($r = 0.927$; $r^2 = 0.86$, i.e. 86% of variance was explained). VL seemed to have the strongest correlation with the overall performance on individual presentation. It was much greater than with the other sub-constructs, i.e. ideas, organization and PD.

In writing, the correlation between 'vocabulary and language' and 'ideas' was high ($r = 0.870$; $r^2 = 0.76$, i.e. 76% of variance was explained). 'Ideas' had the strongest influence on the overall written performance among the other sub-constructs, i.e. 'organization', 'vocabulary and language' and 'features'.

Table 1.1. Correlation of overall ratings and between ratings of sub -constructs within speaking (N = 173)

Speaking – Individual Presentation	Ideas	Organization	Vocabulary & Language Patterns	Pronunciation & Delivery
Overall	0.944	0.835	0.962	0.952
Ideas	…	0.820	0.906	0.895
Organization	…	…	0.803	0.809
Vocabulary & Language Patterns	…	…	…	0.927

Table 1.2. Correlation of overall ratings and between ratings of sub-constructs within writing (N = 173)

Writing	Ideas	Organization	Vocabulary & Language Patterns
Overall	0.964	0.879	0.931
Ideas	…	0.802	0.870
Organization	…	…	0.748

1.2 Correlation of ratings overall and between ratings of sub-constructs across two macro skills

The results in Table 1.3 show moderate correlations ($r = 0.663$; $r^2 = 0.44$, i.e. 44% of variance was explained) between 'ideas and organization'. However, there was a high correlation ($r = 0.74$; $r^2 = 0.55$, i.e. 55% of variance was explained) between 'vocabulary and language patterns' across the two macro skills.

The 'moderate' correlation levels in 'ideas and organization' indicated that transference of organization and ideas between spoken and written modes was probably limited. Separate teaching strategies are thus probably required for ideas and organization in each macro skill (i.e. speaking and writing).

The 'high' correlation levels in VL showed that transference between spoken and written skills was in place and that vocabulary and language patterns learned in speaking could be transferred to writing and so vice versa. Teaching strategies should try to take advantage of this transference.

The disattenuated estimates (figures in brackets in Table 1.3) of the correlations of overall ratings and between ratings of sub-constructs across two skills were calculated using the formula mentioned in 'Methods and Analysis' section. The weakening effects due to measurement error were not observed in 'overall ratings' and 'ideas & organization'. For 'language', the disattenusated value of this estimate was greater than 1.00, indicating that measurement error was not randomly distributed. It may also be due to the lower reliability of language in speaking (i.e. 0.46) as compared to that in writing (i.e. 0.84).

Table 1.3. Correlation of overall ratings and between ratings of sub-constructs across two macro skills (N = 173)

Writing \ Speaking	Overall	Ideas & Organization	Vocabulary & Language Patterns
Overall	0.727 (0.769)	0.710	0.709
Ideas & Organization	0.678	0.663 (0.706)	0.665
Vocabulary & Language Patterns	0.737	0.724	0.735 (1.00)

Note: Figures in brackets refer to the disattenuated estimates of the correlations.

RQ2. To what extent do the rated and counted features of the sub-constructs within a skill and across two skills correlate?

2.1 Correlation between rated and counted features of spoken performances

In Table 2.1, in speaking, 'types' seemed to have the strongest correlation (r = 0.82) with overall performance ratings, followed by 'pronunciation accuracy index' (r = 0.80). 'Types', 'pronunciation' and 'grammar' were the strongest predictors of students' overall spoken performance ratings.

The VQM, 'types' and 'grammar', seemed to be the most powerful predictors of 'vocabulary and language features' (VL). The other VQM, except for TTR, only had 'medium' to 'high' levels of correlations.

The counted features of 'pronunciation and delivery' (PD) i.e. 'pronunciation accuracy index' and 'fluency index' had high correlation levels with all rated features, showing that 'pronunciation' and 'fluency' had a very strong influence on overall spoken performance ratings. The ' pronunciation accuracy index' also seemed to be a very powerful predictor of PD ratings (0.73) and an even more powerful predictor of ratings for ideas (0.79).

Table 2.1. Correlation between rated and counted features of spoken performances

Speaking	Ideas	Organization	Language		Vocabulary				Pronunciation & Fluency	
Counted / Rated	# T Units	Cohesive Devices	T Unit Index	Grammar	Types	Tokens	Type-Token Ratio (TTR)	D	Pronunciation Accuracy	Fluency
Overall	0.67	0.71	0.73	0.79	0.82	0.78	*-0.43*	0.56	0.80	0.78
Ideas	0.65	0.71	0.73	0.79	0.81	0.78	*-0.42*	0.54	0.79	0.78
Organization	0.58	0.61	0.61	0.69	0.72	0.68	*-0.35*	0.54	0.70	0.69
Vocabulary & Language Patterns	0.64	0.68	0.71	0.75	0.79	0.75	*-0.39*	0.56	0.76	0.75
Pronunciation & Delivery	0.62	0.66	0.67	0.72	0.74	0.72	*-0.41*	0.50	0.73	0.71

2.2 Correlations between rated and counted features of written performances

In Table 2.2, 'grammar accuracy index' and 'types' had the strongest correlation (r = 0.73) with the overall written performance rating. The results also show that the counted features of 'language' were a stronger predictor of fair average (FA) ratings of 'ideas' (0.53-0.75) than FA ratings of 'language' itself (0.52-0.66).

The correlation between the VQM, number of T units/T unit index and all rated features (i.e. overall FA ratings, FA ratings of ideas, organization and language) was medium (0.52-0.66). In 'organization', the correlation between the VQM 'cohesive device index' and the FA rating of 'organization' was only 0.47 (on the margins between 'low' and 'medium'). In 'vocabulary', except for type-ration token (TTR), the correlation between the VQM and FA ratings of 'vocabulary and language (VL)' was medium (0.40-0.75). The VQM of 'grammar' and 'types', seemed to be strong predictors of 'VL' (0.68), but they were even stronger predictors of 'ideas' (0.75). 'D' derived by *vocd* only had a 'medium' level (0.52) of correlation with the FA ratings of language.

Table 2.2. Correlations between rated and counted features of written performances

Writing	Ideas	Organization	Language					Vocabulary			
Counted / Rated	# T Units	Cohesive Devices	T Unit Index	Grammar	Spelling	Punctuation	Capitali-zation	Types	Tokens	Type-Token Ratio (TTR)	D
Overall	0.63	0.49	0.60	0.73	0.69	0.69	0.70	0.73	0.68	*0.03*	0.50
Ideas	0.66	0.52	0.63	0.75	0.71	0.71	0.72	0.75	0.71	*0.01*	0.49
Organization	0.57	0.47	0.53	0.66	0.63	0.63	0.64	0.64	0.63	*-0.04*	0.40
Vocabulary & Language Patterns	0.52	0.45	0.52	0.66	0.61	0.61	0.62	0.68	0.60	*0.10*	0.52

2.3 Correlations between counted features of spoken and written performances

In Table 2.3, the correlation levels of the 'vocabulary' VQM (0.50-0.64), except for TTR, between two skills were the highest, followed by 'language' VQM (0.43-0.55), 'ideas' VQM (0.44) and 'organisation' VQM (0.26). This means that 'vocabulary and language' is the highest predictor among the sub-constructs, followed by 'ideas' and 'organization'.

In Table 2.3, 'types' gave the highest correlation figure ($r = 0.64$; $r^2 = 0.41$, i.e. 41% of variance was explained). The second highest was 'grammatical accuracy index' ($r = 0.55$; $r^2 = 0.30$, i.e. 30% of variance was explained) followed by 'D' 'lexical diversity' ($r = 0.54$; $r^2 = 0.29$, i.e. 29% of variance was explained).

Type-token ratio (TTR) and 'D' are supposed to give figures for vocabulary richness. In this study, 'D' gave a higher correlation (0.54; $r^2 = 0.29$, i.e. 29% of variance was explained) between the spoken and written performances than TTR which had a very low correlation level ($r = 0.16$; $r^2 = 0.03$, i.e. 3% of variance was explained). This result supports the arguments of Iwashita, et al., (2001), Richards (1987) and Vermeer (2000) who are concerned about the use of ratio measures like TTR in short performances. They argue that if the performance is short, then the difference in the volume of clauses and T-units between the high ability levels and the lower ability levels will be suppressed.

Table 2.3. Correlations between counted features of spoken and written performances

Criteria	VQM Covariance	Correlations 'r'	r^2	Relationship
Ideas	No. of T-Units	0.44	19%	Medium
Organization	Cohesive Device Index	0.26	7%	Low
Language	T-Unit Index (Syntactic Complexity)	0.43	18%	Medium
	Grammatical Accuracy Index	0.55	30%	Medium
Vocabulary	Types (Number of different words)	0.64	41%	Medium
	Tokens (Number of words)	0.50	25%	Medium
	Type-Token Ratio (Lexical Variation)	0.16	3%	Low
	D (Lexical Diversity)	0.54	29%	Medium

RQ3. What can we deduce from comparing student performances in speaking and writing tasks?

Three sets of speaking (individual presentation) and writing performances of the same students were selected, i.e. best, average and weak, from the sample ($N = 173$). Exemplars of average student performances in both speaking and writing (of the same student) are shown as follows.

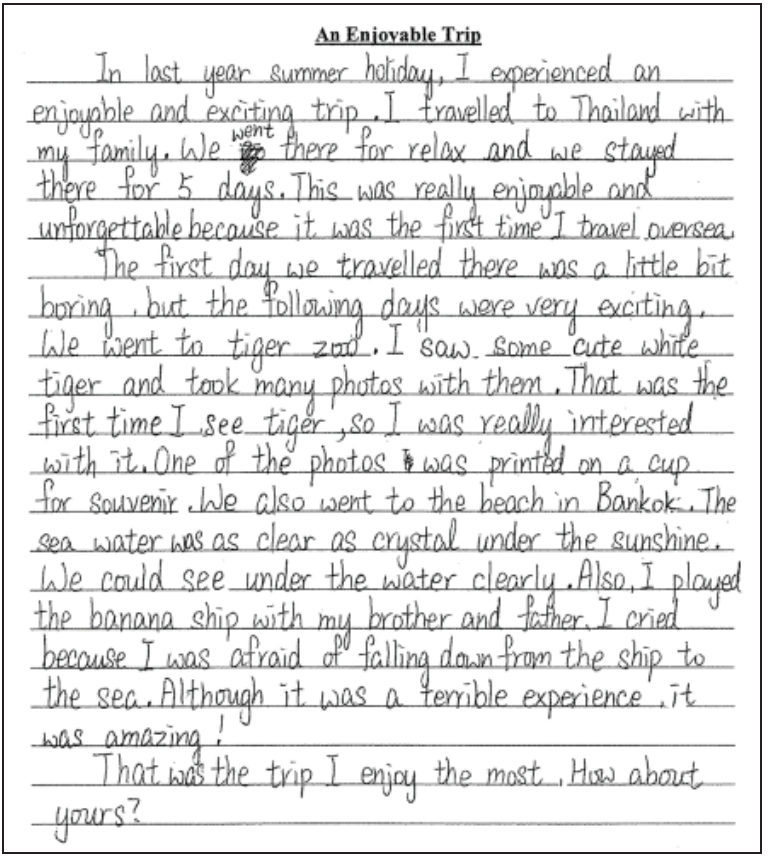
3.1 Characteristics of average student performances

In regards to the average performance on speaking and writing for the same student (see Student Exemplars 3.1 and 3.2 for actual student work and Table 3.1 for annotations), it is found that in terms of content, students can adequately express ideas that are relevant to the topic. At times, they are able to provide some details to expand their ideas in their written task. In the organization of a discourse, use of explicit cohesive devices to link ideas is found in both speaking and writing. Furthermore, slightly more varied language patterns are found in the written tasks than in the spoken tasks. Though complex sentences are found in both skills, they are not used as accurately in spoken mode as they are in written mode. Passive voice is occasionally found in written mode but not in spoken mode. Students' lexical diversity (i.e. range of vocabulary) is slightly higher in the written than in the spoken mode.

Student Exemplar 3.1. Individual Presentation (Average Student Performance)

Good morning everybody. Today I am going to talk… to tell you about a school activity I like… I really like. The school activity I like the most is the St John week. It is… um… it.. eh.. the St. John week hold every years and during that week.. eh … there are many activities about.. eh… the knowledge of eh… the first-aid and it's hold every years…and also there are many activities we can do on that day.. that week. On the first day, there will be some members of St. John will march eh.. in the playground showing us eh… what they would do during the meeting. And also there are eh… questionnaire that is full of the simple first-aid question. Also they will have a demonstration on how to do the first aid eh.. eh.. in some dangerous situation. I like this week because em.. it is really meaningful and.. and we can learn… em many things during this week, also.. em it is also very useful in our daily life. That's the end of my presentation. Thank you.

Student Exemplar 3.2. Writing (Average Student Performance)



An Enjoyable Trip

In last year summer holiday, I experienced an enjoyable and exciting trip. I travelled to Thailand with my family. We went there for relax and we stayed there for 5 days. This was really enjoyable and unforgettable because it was the first time I travel oversea.

The first day we travelled there was a little bit boring, but the following days were very exciting. We went to tiger zoo. I saw some cute white tiger and took many photos with them. That was the first time I see tiger, so I was really interested with it. One of the photos I was printed on a cup for souvenir. We also went to the beach in Bankok. The sea water was as clear as crystal under the sunshine. We could see under the water clearly. Also, I played the banana ship with my brother and father. I cried because I was afraid of falling down from the ship to the sea. Although it was a terrible experience, it was amazing!

That was the trip I enjoy the most. How about yours?

Table 3.1 Average student performances in speaking and writing in the study sample (N = 173)

Individual Presentation – School Activity I Like	**Topic –An Enjoyable Trip**
Speaking (Ideas)	**Writing (Content)**
• Express adequate ideas that are relevant to the topic	• Most or all relevant content; relates ideas to the topic with details
◦ Activity – St John('s) Week - when: hold* every year - where: members of St John('s) march in the playground - what: many activities about knowledge of first-aid, show what would do during the meeting, questionnaire (quiz), demonstration on how to do the first aid in some situation(s) - why: meaningful and learn many things, useful in our daily life	◦ Paragraph 1 – Introduction to the topic (experienced an enjoyable trip in Thailand) - Ideas: travelled to Thailand for 5 (five) days - Explain: enjoyable and unforgettable because it was the first time I travel* overseas ◦ Paragraph 2 – What I did in Thailand - Ideas – the first day was a little boring but the following days were very exciting - Details: tiger zoo – saw some cute white tiger(s) and took many photos with them, one of the photos was printed on a cup for souvenir* - Explain: very exciting because that was the first time that I saw (a) tiger - Details: beach, sea water clear as crystal, played banana ship, cried - Explain: afraid of falling down from the ship ◦ Paragraph 3 – Ending - Inform: the trip I enjoy the most
Speaking (Organization)	**Writing (Organization)**
• Use of simple connectives and sequencers (and, but, first, then, etc)	• Paragraphs generally developed based on prompts • Use of simple connectives and sequencers (and, but, first, then, etc)
◦ Use of cohesive devices - enumerative linking adverbs: e.g. on the first day - additive: e.g. also, and - result: e.g. so, because - following pronoun references: e.g. it	◦ Use of cohesive devices - enumerative linking adverbs: e.g. the first day, the following days - additive: e.g. also, and - result: because, so - concession: although, but - following pronoun references: e.g. that

Speaking (Vocabulary and Language Patterns)	Writing (Vocabulary and Language Patterns)
• Use varied language patterns but with errors that may impede communication (errors with *) • Use varied and appropriate language patterns	• Use familiar vocabulary • Use varied language patterns but lack clarity • Write with some errors in grammar and spelling that does not affect meaning
◦ Complex sentences - The school activity I like the most is the* St John* Week. - There are* questionnaire that is full of simple first-aid question*. - There are many activities about the knowledge of the* first-aid and it's hold* every years*. - There will be* some members of St John* will march in the playground showing us what they would do during the* meeting* (wrong structure) - They will have a demonstration on how to do the* firstaid in some dangerous situation*. - I like this week because… this is really meaningful and… and we can learn… many things during this week. ◦ Modal - …we can learn	◦ Complex sentences - This was really enjoyable and unforgettable because it was the first time I travel* oversea*. - Although it was a terrible experience, it was amazing! ◦ Compound sentences - …, but the following days were very exciting - That was the first time I see* tiger*, so I was really interested with* it. - We went there for relax* and we stayed there for 5 (five) days. ◦ Passive Voice - One of the photos was printed on a cup for souvenir*. ◦ Others - The sea water was as clear as crystal… - How about yours?
◦ Familiar vocabulary - verbs: like, hold*, march, learn - adjectives: simple, dangerous, meaningful, useful, daily - nouns: knowledge, first-aid, members, playground, questionnaire, demonstration, situation, life	◦ Appropriate vocabulary - verbs: experienced, travelled, stayed, printed, relax, cried - adjectives: enjoyable, exciting, boring, cute, interested, clear, terrible, amazing - adverbs: really, clearly, oversea(s)* - nouns: souvenir, beach, crystal, sunshine, experience
Mistakes found in speaking but not in writing - use of passive voice: St John* week hold* every years* - Structures: there will be some members of St John will march in the playground	

Note: * sentences/phrases contain error

Main Findings

Transference from writing to speaking

The 'high' correlation levels in 'vocabulary and language patterns' ratings showed that transference between spoken and written skills is probably in place and that vocabulary and language patterns learned in writing can be transferred to speaking and so vice versa. However, transference from writing to speaking seems more likely given the prioritization of the written mode in the Hong Kong school system and the fact that Hong Kong students usually do not have much chance to practise oral English outside school.

*Predictors of s*tudents' *performances*

The counted features, 'grammatical accuracy index' and 'types (number of different words)' used in this study were the strongest predictors of students' overall written and spoken performance ratings. In writing, 'cohesive device index' was not a strong predictor showing students' organization ratings. This may be because counting of cohesive devices in this study was mainly based on the students' use of explicit cohesive devices. However, students at the high end of the ability range in writing were able to create coherent links using topic sentences and to follow each idea with another related idea (lexical cohesion).

In speaking, the 'pronunciation accuracy index' seemed to be a powerful predictor of 'pronunciation and delivery' ratings and an even more powerful predictor of 'ideas' ratings. It was followed closely in predictive power by the 'fluency index'. One explanation for the effect of fluency on ideas ratings is that a hesitant delivery robs the speaker of paralinguistic tools (such as volume, pitch, pause length and speed of utterance) with which to mark ideas and their relative importance within a text. This in turn results in a poor ideas rating. Similarly, poor pronunciation robs the speaker of the ability to distinguish time, action status (e.g. in progress or completed) and plurality in a discourse, thus reducing coherence. It also can result in confusion between similar sounding words. This also reduces coherence, as perceived by listeners. Both of these reductions in coherence could result in a poor 'organization' rating. This apparent effect of the 'pronunciation accuracy index' or the 'fluency index' on

'vocabulary and language patterns' (VL) could be due to the poor impressions made on raters by hesitant delivery and mispronunciations. Such poor impressions could easily result in a lower VL rating.

Covariance between VQM of productive skills

'Types' showed the 'moderate' level across two macro skills, followed by 'grammatical accuracy index' and 'D' (for lexical diversity). Except for type-token ration (TTR), the correlation levels of 'vocabulary' VQM between the two skills were the highest, followed by those of 'language' and 'ideas'. 'Organization' in terms of the VQM 'cohesive device index' gave only 'low' correlations between VQM values for speaking and writing. This means that 'language' and 'vocabulary' had a higher chance where transference from writing to speaking occurs and so vice versa. These results echoed the aforementioned findings regarding the rated features of the sub-constructs across two macro skills.

Students who have problems in written grammar may also have problems in spoken grammar and students' vocabulary power in speaking to a certain extent predicts their vocabulary power in writing, or the other way round. However, the 'low' correlation of 'cohesive device index' values between speaking and writing suggests that transference in 'organization' between the two macro skills is very limited and separate teaching of coherence for speaking and writing is necessary.

Further Development

Based on the aforementioned findings, it was suggested that a computer corpus-based identification and classification system be developed for students' common errors. The system, *Writing ePlatform*, is designed to enhance students' self-learning and to support learning and teaching. The objectives are as follows:

(a) To further analyse students' problems in language output and learning difficulties based on a research-proven framework for a corpus-based human assisted error identification and classification system.

(b) To verify the learning needs of three groups of students (high, middle and low performing groups) based on the framework mentioned in (a), where prompts are given to students when their writing contains errors identifiable and classifiable.

The functions of the *Writing ePlatform* will:

(1) assist with constructing the lexico-grammatical and discoursal/rhetorical knowledge of the target language and the skills required to access and apply that language;
(2) encourage reflection and metacognition, where students are encouraged in developing independent learning and self-confidence; and
(3) develop 'cognitive apprenticeship', where coaching and modeling occur, and where scaffolding is provided to support language learning.

Conclusion

The findings of this study concur with those of Banerjee et al., (2007, pp. 245-246) who suggest that a more realistic pursuit would be to look for the ideal group of measures that, when applied together, produce a learner language profile that could reliably be classified as being at a given level in a predetermined scale. For this reason, the study of syntax in Hong Kong's students at Key Stage 3 (KS3) requires 'home grown' indices – which takes into account the order in which Hong Kong KS3 students acquire English structures.

The development of *Writing ePlatform* makes possible for the study of syntax primarily at KS4. This kind of research work is also of significance to the development of English learning across different key stages. While the order of acquisition serves as a milestone for students at the end of KS3, showing where the students are after they have completed nine-year basic education and giving indications to student learning at Key Stage 2 and Key Stage 4, such research work can be explored at the other key stages to gain insight into the stage-related cognitive and linguistic strategies needed in developing speaking and writing competency.

References

Banerjee, J., Franceschina, F., & Smith, A. M. (2007). *IELTS Research Reports Volume 7*. IELTS Australia and British Council.

Cheung, K. M. (2010). *Reliability and Validity in Practice: Hong Kong's Key Stage 3 Oral Assessment*. Unpublished PhD Thesis. Macquarie University, Australia.

Crookes, G. (1989). Planning and interlanguage variation. *Studies in Second Language Acquisition, 11*, 367-384.

Florez, V. & Hadaway, N. L. (1987, January). Relationship of oral language proficiency and writing behaviours of secondary second language learners. Paper presented at the Southwest Regional Conference of the International Reading Association. (15th, Phoenix, AZ, January 22 – 24, 1987).

Foucault, M. (1974). *Power/knowledge: Selected interviews and other writings*. New York: Pantheon Books.

Hawkey, R., & Barker, F. (2004). Developing a common scale for the assessment of writing. *Assessing Writing, 9*, 122-59.

Heatley, A., Nation, I. S. P., & Coxhead, A. (2002). RANGE and FREQUENCY programs. Retrieved from http://www.vuw.ac.nz/lals/staff/Paul_Nation on 3 March 2008.

Hubert, M. D. (2008) *The relationship between writing and speaking in the U.S. university Spanish language classroom.* Unpublished Thesis. Purdue University Library.

Iwashita, N., McNamara, T., & Elder, C. (2001). Can we predict task difficulty in an oral proficiency test? Exploring the potential of an information processing approach to task design. *Language Learning, 21*, 401-436.

Kim, A. (2000). A comparison of oral & writing development in a second language college student. *Research and Teaching in Developmental Education, 16*(2). 73-82.

Linacre, J. M. (1991-2008). *FACETS. Version 4.64* (computer program).

Malvern, D., Richards, B., Chipere, N., & Duran, P. (2004). Lexical diversity and language development: quantification and assessment. Palgrave, Macmillan.

Ortega, L. (1999). Planning and focus on form in L2 oral performance. *Studies in Second Language Acquisition, 21*, 109-48.

Richards, B. (1987). Type/token ratios: What do they really tell us? *Journal of Child Language, 14*, 201-09.

Skehan, P., & Foster, P. (1999). The influence of task structure and processing conditions on narrative retellings. *Language Learning, 49*, 93-120.

Vermeer, A. (2000). Coming to grips with lexical richness in spontaneous speech data. *Language Testing, 17,* 65-84.

Zhu, X. H. (2007, March). What do we know about the relationship between speaking and writing in college-level ESL students? *US-China Foreign Language*, *Volume 5*, No.3 (Serial No.42). USA.

Authors' e-mail: Amy Cheung kmcheung@hkeaa.edu.hk
Flora Leung scdobcae@edb.gov.hk