TẠP CHÍ NGHIÊN CỨU NƯỚC NGOÀI
VNU JOURNAL OF FOREIGN STUDIES
ISSN 2525-2445

Xuất bản 01 kỳ/02 tháng

Giấy phép hoạt động báo chí in
Số 550/GP-BTTTT ngày 09/12/2016
của Bộ Thông tin và Truyền thông

Tạp chí *Nghiên cứu Nước ngoài*, Tầng 3, Nhà A1, Trường Đại học Ngoại ngữ, Đại học Quốc gia Hà Nội, Phạm Văn Đồng, Cầu Giấy, Hà Nội, Việt Nam

*** ĐT.: (84-24) 62532956 * Email: tapchincnn@gmail.com / tapchincnn@vnu.edu.vn**
*** Website: https://js.vnu.edu.vn/FS/**

VNU JOURNAL OF FOREIGN STUDIES
Vol.36, No.4, 2020

CONTENTS

TẠP CHÍ NGHIÊN CỨU NƯỚC NGOÀI
Tập 36, Số 4, 2020

MỤC LỤC

NGHIÊN CỨU

THE "GENERAL SCHOOL EDUCATION INTRODUCTORY ENGLISH CURRICULUM FOR GRADE 1 AND GRADE 2": A NEW INNOVATION IN FOREIGN LANGUAGE GENERAL SCHOOL EDUCATION IN VIETNAM

Hoang Van Van*

VNU University of Languages and International Studies,
Pham Van Dong, Cau Giay, Hanoi, Vietnam

Received 05 October 2019
Revised 12 March 2020; Accepted 20 July 2020

Abstract: This paper is concerned with a new innovation in foreign language general school education in Vietnam which has been attracting considerable attention from Vietnamese foreign language educators, foreign language teachers, and the general public – the *General School Education Introductory English Curriculum for Grade 1 and Grade 2* issued in 2018 by the Ministry of Education of Vietnam. To highlight this important innovation, the paper attempts to address three main issues: (1) it will examine and clarify some of the most important aspects of the *General School Education Introductory English Curriculum for Grade 1 and Grade 2*; (2) it will present and discuss in some detail the contents in *Tiếng Anh 1* (English 1) – a new English textbook developed in accordance with the *General School Education Introductory English Curriculum for Grade 1 and Grade 2* by Vietnam National Publishing House in collaboration with Springer Nature Macmillan Education; and (3) it will outline and discuss some specific classroom methods and present a new procedure for teaching an activity/task to help teachers and pupils teach and learn English effectively.

Keywords: the *General School Education Introductory English Curriculum for Grade 1 and Grade 2*, *English 1*, effective classroom teaching methods, new procedure for teaching a language activity

1. Introduction

In the present-day world of increasing integration and globalization, English is naturally recognized as having an extremely important role. Starting from a collection of dialects spoken in the southern counties of England, after nearly four centuries the English language has expanded far beyond its nation, and has now become a *lingua franca* in two distinct but related senses: "international language" and "global language" (Halliday, 2017). English has been the most widely used language in literature, art, business, science, technology, media, diplomacy, tourism and many other areas of human activity. Recent studies on the situation of the teaching and learning of English in the world (e.g. Cheshire, 1996; Nunan, 2003; Eurydice, 2005, 2017; Shin and Scrandall, 2015; Hoang Van Van, 2016, 2019) have shown that in many countries and territories around the world a foreign language (principally English) is introduced

* Tel.: 84-946296999, Email: vanhv@vnu.edu.vn; vanhv.sdh@gmail.com

into the primary school, and in some countries and territories English is even introduced into the pre-school (the kindergarten level).

In Vietnam, since *Doi Moi* (Renovation) which was initiated by the Communist Party of Vietnam in 1986, English has become the most important foreign language taught at all levels of learning nationwide. In 2008, the Prime Minister of the Socialist Republic of Vietnam issued Decision 1400 approving the national Project entitled *Đề án dạy và học ngoại ngữ trong hệ thống giáo dục quốc dân giai đoạn 2008-2020 (Teaching and Learning Foreign Languages in the National Education System Period 2008-2020)* which states that at the general school education, English is a compulsory subject taught from grade 3 through to grade 12. Since the end of 2010, the Ministry of Education and Training (MoET) (Bộ Giáo dục và Đào tạo, 2010a, 2012a, 2012b) has issued and piloted nationwide three English language curricula referred to respectively in Vietnamese as *Chương trình tiếng Anh thí điểm tiểu học* (Pilot English Curriculum for Primary Schools in Vietnam), *Chương trình giáo dục phổ thông môn tiếng Anh thí điểm cấp trung học cơ sở* (Pilot English Curriculum for Lower Secondary Schools in Vietnam), *Chương trình giáo dục phổ thông môn tiếng Anh thí điểm cấp trung học phổ thông* (Pilot English Curriculum for Upper Secondary Schools in Vietnam) (for details of these English pilot curricula, see Hoang Van Van, 2018). In 2018, MoET (Bộ Giáo dục và Đào tạo, 2018a) had the three pilot English language curricula combined into one unified English curriculum, and issued it under the title of *Chương trình giáo dục phổ thông: Chương trình môn tiếng Anh* (General School Education Curriculum: English Curriculum). At the same time, in order to meet the ever increasing needs of learning English of early primary children, MoET (Bộ Giáo dục và Đào tạo, 2018b) issued a separate English language curriculum entitled *Chương trình giáo dục phổ thông làm quen tiếng Anh lớp 1 và lớp 2* (General School Education Introductory English Curriculum for Grade 1 and Grade 2). This is perhaps the first optional introductory English curriculum in the world issued at national level. "What are the main features of the *General School Education Introductory English Curriculum for Grade 1 and Grade 2*?", "How are the guidelines in the curriculum transformed into English textbooks for grade 1 and grade 2 children?", and "What possible method(s) should be suggested to teach English effectively to Vietnamese grade 1 and grade 2 pupils?" These questions will be addressed throughout this article. The article is organized around five sections. Following Section one which introduces and outlines the study, Section two examines and highlights some of the most important aspects of the *General School Education Introductory English Curriculum for Grade 1 and Grade 2*. Section three presents and discusses in some detail the contents in *Tiếng Anh 1* (English 1) – a new English textbook developed in accordance with the *General School Education Introductory English Curriculum for Grade 1 and Grade 2* by Vietnam National Publishing House in collaboration with Springer Nature Macmillan Education. Section four outlines some specific classroom methods and presents a new procedure for teaching an activity/task to help teachers teach English to grade 1 and grade 2 children effectively. Finally, Section five summarizes the main points explored in the paper, offering some ideas on how to achieve the goals of the *General School Education Introductory English Curriculum for Grade 1 and Grade 2*.

2. The General School Education Introductory English Curriculum for Grade 1 and Grade 2

The General School Education

Introductory English Curriculum for Grade 1 and Grade 2 (hereafter shortened to *the Curriculum*) was promulgated by the Ministry of Education and Training via Circular No. 32, December 26, 2018. This is a major document all parts of which are central to our discussion, but for the purpose of this paper, we are focusing on four headings: (1) characteristics of *the Curriculum*, (2) principles for designing *the Curriculum*, (3) goal and objectives of *the Curriculum*, and (4) content of *the Curriculum*.

2.1. Characteristics of the Curriculum

The Curriculum is characterized by the following prominent features:

• It is an **optional** curriculum.

• Its aim is to familiarize pupils with very simple and basic or "introductory" English.

• It is a **national level** curriculum: it is issued by the Vietnamese Ministry of Education and Training. Some may notice that some other countries in the world have also introduced English in the early years of the primary level (grade 1 and grade 2), but no country seems to have issued a similar English optional curriculum at the national level as Vietnam.

• It is designed in connection with the *General School Education English Curriculum* – the 10-year curriculum in which English is taught as a compulsory subject from grade 3 through to grade 12 issued by the Ministry of Education and Training via Circular No. 32/2018/-TT-BGDĐT of December 26, 2018 (Bộ Giáo dục và Đào tạo [MoET], 2018b).

2.2. Principles for designing the Curriculum

The design of *the Curriculum* is based on the following principles:

• It should comply with the guidance on optional subjects as defined in the *General School Education Curriculum* promulgated by the Ministry of Education and Training via Circular No. 32/2018/-TT-BGDĐ of December 26, 2018 (Bộ giáo dục và Đào tạo [MoET], 2018a).

• It should be based on the theoretical and practical realities of teaching English as a foreign language, drawing on the achievements in teaching foreign languages to young children in the world as well as in Vietnam, and, in particular, taking into account social practices, economic conditions and cultural traditions of Vietnam, the background diversity of Vietnamese children in terms of regions, environments, and abilities.

• It should be communication-based, taking communicative competence as the goal of the teaching process, and language knowledge as a means to form and develop pupils' communication skills. In the initial stage of learning, it should give priority to oral skills, particularly the listening skill, while paying due attention to initial writing skills such as tracing and then writing alphabetical letters in words.

• It should be organized according to communicative competence development goals, language goals (which are seen as a means to form and develop communication skills), and themes and topics (which are familiar to pupils and allow for recycling and expanding in a spiral circle over the period of two years of learning to develop and strengthen children' communicative competence).

• It should be based on "centre design" principle, taking methods of teaching as the point of departure for setting *the Curriculum's* learning outcomes. Apart from this, it should take into account learners' factors (such as their psychological and physical characteristics, their learning styles, their time of exposure to English language), teachers' factors (such as their qualification, their current classroom teaching methods, their role in guiding pupil's learning), and the diverse teaching and learning conditions across Vietnam.

• One more principle which can be inferred from the design of *the Curriculum* is that it should ensure the continuity of curriculum between it and the *General School Education English Curriculum* (Bộ giáo dục và Đào tạo [MoET], 2018b).

2.3. Goal and objectives of the Curriculum

The overarching goal of *the Curriculum* is to provide Vietnamese grade 1 and grade 2 children with basic "introductory" or "crash-level" English. Through exploring and experiencing introductory English, pupils develop basic English communication skills in contexts suitable to their age level of thinking, their emotional and physiological ability, helping them feel confident when they enter grade 3 where they will learn English as a compulsory subject through to grade 12, and gradually developing their love for the English language as a school subject.

To accomplish this overarching goal, *the Curriculum* sets 12 specific objectives which are stated on behavioural terms. Accordingly, on finishing the programme, pupils can:

• listen and recognize letters in the English alphabet;

• listen and recognize some basic sounds corresponding to the letters in the English alphabet;

• listen and understand numbers from 1 to 20;

• listen and understand simple words and phrases related to pupils' experience and activities in contexts;

• listen, understand, and give nonverbal responses in simple communication activities related to grade 1 and grade 2 children's experience;

• listen and understand simple English classroom instructions;

• listen, understand and give responses in English in simple and familiar question-and-answer situations;

• answer simple questions familiar with pupils' experience at word and very simple sentence level;

• give some simple and familiar English instructions when engaging in classroom communicative activities;

• recognize and read aloud very simple, concrete and familiar words and phrases;

• write some very simple words related to familiar topics;

• form a love for the English language as a school subject.

Based on the goal and objectives, *the Curriculum* outlines the requirements for grade 1 and grade 2 pupils in terms of the four skills of listening, speaking, reading, and writing.

On finishing grade 1, pupils can:

• repeat words, phrases, simple sentences, chants, and songs suitable to grade 1 pupils' level of thinking, their emotional and physiological ability, and their psychosocial skills;

• number 1 to 10, and count them or give answers to simple questions relating to these numbers;

• recognize and name simple and concrete words related to the learned topics in specific contexts of communication;

• respond to questions in English in very simple question-and-answer situations;

• listen and understand nonverbal responses in simple conversations and use simple classroom commands;

• know how to use simple greeting and simple leave-taking expressions.

On finishing grade 2, pupils can:

• repeat simple phrases, simple sentences, chants, and songs suitable to grade 2 pupils' level of thinking, their emotional and physiological ability, and their psychosocial skills;

• recognize and name concrete nouns and verbs related to the learned topics in specific contexts of communication;

• recognize numbers 11 to 20, and count them or give answers to simple questions relating to these numbers;

• listen, understand and give simple feedback in English in similar and very simple question-and-answer situations;

• understand and give very simple classroom instructions which are longer and more complex than those learned in grade 1;

• give some familiar and simple commands when engaging in communicative activities in class, using more words and more complex grammatical structures than those learned in grade 1;

• answer simple questions and give one- or two-word answers in specific and familiar contexts.

2.4. Content of the Curriculum

The content of *the Curriculum* is organized around two main headings: General content and specific content.

2.4.1. General content

The general content of *the Curriculum* consists of three components: language content, language skills, and a list of themes:

Language content consists of three aspects: pronunciation or phonics, vocabulary, and grammatical structures which are respectively specified in *the Curriculum* as follows:

• **Pronunciation**: (Pupils) are introduced to some basic single or monophonic sounds: vowels, consonants (in initial and final word position), some letter(s) realizing the sound(s).

• **Vocabulary**: are introduced to simple words and phrases relating to people, things, and phenomena close to their daily experience in familiar topics and situations. The number of active words to be introduced ranges from 70 to 140.

• **Grammatical structure**: are introduced to some simple grammatical structures, and are expected to be able to use them in familiar and very simple communicative situations.

Language skills include four aspects of speech: listening, speaking, reading, and writing which are respectively stated in behavioural terms as follows:

• **Listening**: (Pupils) can listen, understand, and answer very simple questions (using some very simple appropriate nonverbal expressions or gestures) at word level in simple classroom conversations and in familiar topics or situations; can listen, understand, and follow simple classroom instructions in English; and can understand simple words and phrases close to their experience in familiar contexts or topics within the recommended range of 70-140 words.

• **Speaking**: can repeat simple words, phrases, and sentences; can sing chants and songs related to the topics suitable to the age of grade 1 and grade 2 children; can answer simple questions in familiar topics; can participate in simple communicative activities and language games; and can give some simple English instructions.

• **Reading**: can read simple words and sentences with the help of illustrative images; can listen to words and simple structures and read them along; and can recognize words and understand their meaning in the learned topics within the recommended range of 70-140 words.

• **Writing**: can trace and write single (alphabetical) letters and words (at grade 1); and can write single letters, simple words, phrases, and sentences in specific contexts (at grade 2).

List of Themes. *The Curriculum* recommended a list of 21 general themes (similar to "notions" in David Wilkin's (1976) terminology). This is a useful suggestion for textbook writers as they can base themselves on this list of themes to break it down into smaller and more specific topics for their textbook design. Details are provided in Table 1.

Table 1. List of themes recommended in the Curriculum

1. Colours	8. Daily activities	15. Fun activities
2. Animals	9. Classroom activities	16. Rooms in the house
3. Toys	10. School things	17. Fruit
4. School	11. Parts of the body	18. Food
5. Family	12. Days of the week	19. Emotions
6. Basic shapes	13. Clothes	20. Senses
7. Locations	14. Means of transport	21. Games

2.4.2. Specific content

Based on the list of themes, *the Curriculum* suggests specific contents which grade 1 and grade 2 children are required to learn. These include topics (which are derived from the list of themes), communicative competences which pupils need to learn in terms of the language skills of listening, speaking, reading, and writing, the number of sounds, of letters, of words, and of grammatical structures. Due to limited space, however, these components are not presented here. (For details, readers are referred to Bộ Giáo dục và Đào tạo [MoET], 2018c).

3. English textbooks for grade 1 and grade 2

Textbooks are an almost universal element in any form of teaching. For the majority of school teachers, textbooks serve as a peg on which teachers and students hang their teaching and learning (O'Neill, 1995; Hoang Van Van, 2012). For teachers of English at primary level, textbooks play an even much more important role (Rixon and Papp, 2018). They "provide structure and a syllabus for a programme (curriculum in our terminology); help standardize instruction; maintain quality; provide a variety of learning resources; are sufficient (in the sense of saving teachers' time, enabling them to devote time to teaching rather than material production); provide effective language models and input; and train teachers" (Richards, 2001: 254-55). They "offer a coherent syllabus, satisfactory language control, motivating texts, tapes and other accessories such as videotapes, CD-ROMS, extra resource material, and useful weblinks. (…). They come with detailed teacher's guides which not only provide procedure for the lesson in the student's book, but also offer suggestions and alternatives, extra activities and resources (Hammer, 2005: 304; 2017: 152-53). These are perhaps the reasons why when asked to provide comments about a curriculum, many EFL teachers often give responses in relation to a textbook without knowing that a textbook is just a realization of the curriculum. At present, Vietnam is exercising the policy of "one curriculum, multiple textbooks", and there are actually several sets of English textbooks in use in primary schools. In what follows, I shall be concerned only with the set of English textbooks developed for grade 1 and grade 2 children by authors of Vietnam Education Publishing House (VEPH) in collaboration with those of Springer Nature Macmillan Education (hereafter shortened to the VEPH-Macmillan set of textbooks). The VEPH-Macmillan set of textbooks has been trialled in a number of primary schools in Vietnam and has been well received for its pedagogy and resource design; its components comprising *Sách học sinh* (Student's Books), *Sách giáo viên* (Teacher's Books), *Sách bài tập* (Workbooks) CD-ROMS, and other supplementary learning resources.

Given that the time allocated for each of the grades in *the Curriculum* is very limited: 2 periods/per week, and 70 periods/per school year; given that English for grade 1 and grade 2 is an optional subject; and given that the aim of *the Curriculum* is to familiarize the children with very basic "introductory" English, the following questions were raised for consideration when developing the set of textbooks:

- How much language content (pronunciation, vocabulary, and structure) should be selected for the design of the textbooks to accomplish *the Curriculum's* goal?
- What level of English in terms of language skills is needed in the design of the textbooks to accomplish *the Curriculum's* goals?
- How should the language content and language skills be integrated into the topics of the textbooks to accomplish *the Curriculum's* goal?
- How much intercultural content (Vietnamese, English-speaking, and international culture) is needed, and how should it be integrated into the design of the textbooks to accomplish *the Curriculum's* goal?
- How should the textbooks be designed to help teachers and pupils achieve *the Curriculum's* goal?

As *Tiếng Anh 2* has not yet come out, we will address these questions in relation to *Tiếng Anh 1 – Sách học sinh* (English 1 – Student's Book), *Tiếng Anh 1 - Sách giáo viên* (English 1 – Teacher's Book), and *Tiếng Anh 1 - Sách bài tập* (English 1 – Workbook) – an intercultural collaboration product co-developed by VEPH and Springer Nature Macmillan Education (Hoang Van Van et al., 2019a, 2019b, 2019c).

3.1. English 1 - Student's Book

Aim of writing

English 1 - Student's Book is designed in accordance with the *General School Education Introductory English Curriculum for Grade 1 and Grade 2*. Its aim is to provide grade 1 children with basic English in terms of knowledge and skills so that they will be confident to continue learning English at grade 2 through to grade 12.

Principles of writing

The design of English 1 - Student's Book is based on the following principles:

- It should be designed in accordance with the *General School Education Introductory English Curriculum for Grade 1 and Grade 2* in terms of time frame, content of teaching, method of teaching, and method of testing;
- It should place English in relation to other subjects offered in grade 1, especially the subject of Vietnamese (children's mother tongue), taking this as a basis for selecting, grading, and sequencing the content (linguistic knowledge and

communication skills) of the textbook;

• It should help grade 1 pupils familiarize with "introductory English" in the proper sense of the term at the beginning stage of the primary education;

• It should ensure the continuity of *Tiếng Anh 1* (English 1) with *Tiếng Anh 2* (English 2), and with the rest of the 10-year English textbook series from grade 3 through to grade 12 in the Vietnamese general school education;

• It should balance intercultural features in its design so as to initially familiarize children with the world's cultural diversity.

Approach to writing English 1 – Student's Book

The main approach to writing *English 1 – Student's Book* is communication, taking as centre pupils' learning activities, laying emphasis on pupils' formation of English communication skills through simple interactions in contexts, helping them acquire oral English in the earlier stage of language learning. At the start of the course, particular emphasis is placed on listening and speaking (through repeating, chanting, singing), with due attention being paid to pupils' ability to trace and then write alphabetical letters in words.

The main approach to organizing the contents of *English 1 – Student's Book* is audio-lingual and audio-visual presented in words and simple structures context. The audio-lingual and audio-visual materials and structural patterns are selected, graded, and sequenced in accordance with the selected topics and are presented in colours to facilitate pupils' "absorption" of the language.

Content of English 1 – Student's Book

Linguistic content: *English 1 – Student's Book* focuses on:

Pronunciation:	16 single sounds, including 11 consonants and 5 vowels: *b, c, a, d, i, e, g, h, o, m, u, l, n, t, f, w*.
Vocabulary:	64 basic productive words, including verbs, nouns (singular and plural), pronouns, numbers, adjectives, etc. These are words indicating objects, phenomena, and events relating to situations and topics familiar to pupils' experience.
Structure:	Simple and basic sentence structures with the verb *be,* the modal verb *can*, and with other verbs familiar to pupils' experience such as *have*, *run*, *see*, *touch* used in the present simple, the present continuous, and the imperative mood.

Language skills: Drawing on *the Curriculum's* guidelines, the textbook sets the following language skill requirements. On finishing grade 1, pupils can:

Listening:	listen and recognize letters corresponding to 16 selected single sounds; listen and repeat words and phrases, and words and phrases in simple chants and songs; listen, understand, and give nonverbal feedback in simple communicative situations; listen, understand, and give very simple classroom instructions; listen and understand words and phrases related to familiar topics; listen, recognize, and pronounce learned vowels and initial and final consonants; listen, understand, and count numbers from 1 to 10.
Speaking	ask and answer simple questions, using words or phrases in very simple sentences about school things, toys, colours, numbers, names of family members, etc.; name people and simple concrete objects with the help of images; sing chants and songs to consolidate pronunciation, stress, rhythm, and intonation;

Reading:	read aloud words and simple phrases; sing simple chants; sing simple songs; read and understand very simple stories.
Writing:	trace alphabetical letters; trace alphabetical letters in words; write alphabetical letters in words; write words in phrases.

Structure of English 1 – Student's Book

English 1 – Student's Book includes **16 topics/units**, **4 fun time sections,** and **4 review units**. The 16 topics/units have been selected to provide relevant topics that young Vietnamese children can relate to and engage with, both in and outside of classroom. Vietnamese characters (e.g. Ba, Nam, Hoa, Nam, etc.) in the textbook provide children a sense of familiarity and friendliness, and support them in exploring these topics in English. And overseas characters (e.g. Bill, Ben, Lucy, Phil, Sue, etc.) offer pupils new and different insights, and encourage their interest in discovering more of the wider world. Specific design of the 14 units, 4 fun times, and 4 reviews in *English 1 – Student's Book* is displayed in Table 2.

Table 2. Structure of English 1 – Student's Book

Unit 1	In the school playground	Unit 9	In the shop
Unit 2	In the dining room	Unit 10	At the zoo
	Fun time 1		Fun time 3
Unit 3	At the street market	Unit 11	At the bus stop
Unit 4	In the bedroom	Unit 12	At the lake
	Review 1		Review 3
Unit 5	At the fish and chip shop	Unit 13	In the school canteen
Unit 6	In the classroom	Unit 14	In the toy shop
	Fun time 2		Fun time 4
Unit 7	In the garden	Unit 15	At the football match
Unit 8	In the park	Unit 16	At home
	Review 2		Review 4

Structure of a unit in English 1 – Student's Book

Each unit in *English 1 – Student's Book* contains three individual lessons. Each lesson provides material for one teaching period (equating to thirty five minutes), in which language content and language skills are developed through warm-up activities, presentation, practice, application, and self-assessment/check.

Each unit in *English 1 – Student's Book* has a logical progression of activities, and clear and concise instructions which guide children through to their completion. These activities also help develop coordination, critical thinking and cooperative learning skills, as pupils learn to understand and use English in both its spoken and written forms.

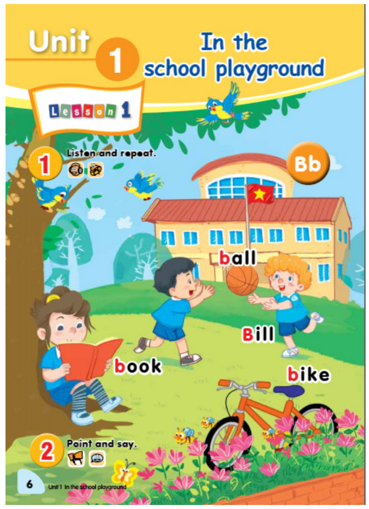
Phonics is presented systematically throughout the units, helping pupils relate sounds to letters (symbols) in words, at the same time initially familiarizing them with word stress, rhythm and intonation in English through singing short chants and short songs.

The structure of a unit and the activities of each of the three lessons in the unit are represented in some detail below.

Lesson 1

Lesson 1 aims to familiarize pupils with the topic of the unit. Opening the *Student's Book* on Page 6, for example, pupils can see the number of the unit (Unit 1) and its title (In the school playground). Under these headings is a whole-page picture of a school scene setting the context for the lesson. In the picture pupils can see two boys playing with a ball, a girl reading a book under a big tree, and a bike standing next to a bed of flowers. Pupils can also see four words with the initial letter (in upper or lower case) "B/b" highlighted in red: **ball**, **Bill, book,** and **bike**. Lesson 1 consists of two activities or tasks: **Listen and repeat** and **Point and say**.

The **Listen and Repeat** activity is intended to introduce pupils to the sound /b/ represented by the upper case letter B or the lower case b in the picture and the four words presented in the context of the picture (from top): **ball**, **Bill, book,** and **bike**. Pupils are required to listen to the target sound and repeat it in in the words.



The **Point and say** activity aims to help pupils pronounce correctly the target sound in the four-word context. By pointing at the letter highlighted in red and say the words in which the target sound occurs, pupils are initially familiarized with the sound-symbol association, and the meaning of the words in context.

Lesson 2

Lesson 2 aims to strengthen the contents pupils have learnt in Lesson 1. It includes three activities: **Listen and chant**, **Listen and tick**, and **Look and trace**/**Look, trace and write**.

The **Listen and chant** activity is intended to strengthen pupils' pronunciation accuracy of the target sound /b/ in association with the letter B/b by listening first and then chant. Through this activity, pupils can also be familiarized with English stress, rhythm and intonation in authentic context.

The **Listen and tick** activity is intended to initially introduce pupils to listening skill. They are given two pictures; each contains two objects (representing two words they have learned). Pupils listen to the recording and give a tick to the correct answer.

The **Look and trace**/**Look, trace and write** activity aims to develop pupils' ability to write alphabetical letters in words. It is designed into three separate stages to suit pupils' writing ability development: **Look and trace** is designed for the first stage (semester 1 of grade 1); **Look, trace and write**, for the second stage (semester 2 of grade 1); and **Look and write**, for the third stage (both semester 1 and semester 2 of grade 2).

Lesson 3

Lesson 3 aims to familiarize pupils with simple interactive English. It consists of three activities: **Listen and repeat**, **Let's talk**, and **Let's sing**.

The **Listen and repeat** activity aims to familiarize pupils with simple "seeded" forms of greeting (greeting and introducing oneself), and leave-taking (saying good bye to someone) in English. Pupils listen to the recording and repeat these forms (e.g. *Hi, I'm Bill* and *Bye, Bill*).

The **Let's talk** activity comes after the **Listen and repeat**. It aims to initially develop pupils' ability to speak English. Pupils use the same "seeded" forms of greeting and leave-taking; fill the blank the name of someone they have chosen at will and practise greeting and leave-taking with each other.

The **Let's sing** activity consists of a very simple English song. Its aim is to help pupils consolidate the sound-symbol association, the words, the phrases, and the structures they have been introduced in the previous seven activities. At the same time, this activity is intended to further consolidate pupils' pronunciation of the learned sound, and to have a sense of English rhythm and intonation as they sing along with actions such as dancing, jumping, or hopping.

Fun time

The **Fun time** section is designed after pupils have finished two learning units. It consists of two lessons.

Lesson 1 consists of two activities: **Find/Match and circle. Then say** and **Let's play**.

The **Find/Match and circle. Then say** activity aims to help pupils revise the words they have learned. They are given a word search with the pictures representing the learned words around. Their task is to find the words, match them with the pictures, and then say them aloud.

The **Let's play** activity aims to help pupils revise the sound-symbol associations they have learned through playing a game such as "Simon says", "Slap the board", etc.

Lesson 2 is also organized into two activities: **Read and tick. Then colour and say** and **Let's play.**

The **Read and tick. Then colour and say** activity aims to help pupils identify the words they have learned from the pictures, then colour them, and say them aloud.

The **Let's play** activity aims to help pupils revise the structures they have learned through playing a game such as "A happy circle", "Air drawing", etc.

Review

The **Review** unit is designed after pupils have finished four learning units. It aims to help pupils revise the sound-symbol association and the words they have learned through a diversity of activities. A **Review** unit generally consists of seven activities which can be delivered in three lessons.

Lesson 1 is a very simple dialogue between two young English characters: Phil and Sue. Their dialogue is about the words pupils have learned. This lesson is designed into two activities: **Listen and repeat** and **Look and circle.**

The **Listen and repeat** activity requires pupils to listen to the dialogue between Phil and Sue and repeat what they said.

The **Look and circle** activity aims to develop pupils' exploratory ability. Pupils are asked to look at the pictures and find the corresponding words from the list of the words which Phil and Sue said in their dialogue, then circle the correct ones.

Lesson 2 and Lesson 3 aim to initially develop pupils' ability to self-evaluate or self-check their progress. These lessons consist of five activities: **Listen and tick**, **Listen and circle**, **Listen and tick or cross, Read and tick,** and **Colour, write and say.**

The **Listen and tick** activity aims to develop pupils' listening ability. They are asked to listen to the words from the audio CD and tick the correct boxes.

The **Listen and circle** activity aims to develop pupils' listening ability. They are asked to listen to the sounds from the audio CD and circle the correct letters representing the sounds.

The **Listen and tick or cross** activity aims to develop pupils' listening ability. They are asked to listen to the words from the audio CD and tick or cross the correct boxes.

The **Read and tick** activity aims to develop pupils' initial reading ability. They are asked to read the words, look at the pictures, and then tick the correct boxes.

The **Colour, write and say** activity aims to consolidate pupils' alphabet letter writing ability. They are asked to colour the words, write the intended letter in the words, and then say them aloud.

3.2. English 1 - Teacher's Book

The **Introduction** part consists of four sections. Following Section one which provides a brief introduction to *English 1 - Student's Book*, Section two is concerned with a description of a learning unit in *English 1 - Student's Book*. Section three provides guidelines for teaching unit activities. And Section four includes some suggested common classroom instructions and expressions.

The **Guidelines for teaching Units, Fun times, and Reviews** part provides teachers with detailed guidance on how to teach the sixteen learning units, four fun time sections, and four review units. Central to grade 1 teaching methodology is the procedure of **G-I-P-O** (detail of which will be presented in

Section 4.4.2 below). Teachers are advised to use this procedure in teaching and in preparing their lesson plans. For the diversity of teaching and learning contexts, it is recommended that teachers employ the procedure flexibly so as to teach their pupils effectively.

English 1 - Teacher's Book is written in English. In addition to the instructions for individual lessons, it provides teachers with English language support and in-classroom management guidance. It is written in simple English so that teachers can grasp the ideas of the *Student's Book* and use the *Teacher's Book* effectively.

The guidelines suggested in *English 1 - Teacher's Book* are based on the learning-centred approach in an effort to help teachers engage their pupils in interpersonal communication activities, and promote their creative learning potential.

Structure of English 1 - Teacher's Book

English 1 - Teacher's Book is organized into two main parts. Part I is **Introduction,** and Part II – **Guidelines for teaching Units, Fun times, and Reviews**.

In order to maximize access to the content, together with the *Student's Book*, the *Workbook* and other supplementary learning resources, the *Teacher's Book* is offered online. It is hosted on a secure website, with password protection for digital management (for detail, readers are referred to sachmem.vn).

3.3. English 1 - Workbook

English 1 - Workbook provides further reinforcement of the content in *English 1 – Student's Book*. The activities in the *Workbook* are designed to be attractive and engaging so that children can review the learned content in an enjoyable and memorable way. Clear instructions and the systematically- organized format in which these activities are presented ensure that the *Workbook* can be used easily outside the classroom. Apart from these, the *Workbook* contains some allowable extension of knowledge and skills to encourage pupils to learn more English.

The content of the *Workbook* is designed according to the sixteen learning units in the *Student's Book*. A unit in the *Workbook* includes seven activities: (1) **Look and match**, **(2) Listen and circle**, **(3) Look and trace**, (4) **Look and write. Then say**, **(5) Read and match**, (6) **Game**, and (7) **Project**.

The **Look and match** activity aims to help pupils revise the meaning of the words they have learned. They are required to look at the words and match them with the correct pictures.

• The **Listen and circle** activity aims to help pupils reinforce listening skill. They are required to listen to the audio CD and circle the correct words they hear.

• The **Look and trace** activity aims to consolidate pupils' initial writing (tracing) skill. They are required to look at the broken line letters and then trace them.

• The **Look and write. Then say** activity aims to consolidate pupils' writing skill. They are required to write the learned letter in the blank of the learned words, and then say the words aloud.

• The **Read and match** activity aims to consolidate pupils' reading skill. They are required to read the words they have learned in the "seeded" structures, and then match them with the correct pictures.

• The **Game** activity aims to reinforce pupils' vocabulary, develop their initial reading and speaking skills through creative and enjoyable activities such as "Do the puzzle", "Point to the difference. Then say", etc.

• Each unit of the *Workbook* ends with a **Project**. This activity aims to help pupils (i) revise the sound and the letter in the words they have learned from the learning unit, and (ii) to encourage pupils' creativeness in a real and engaging context such as "Trace the apple, the bag, the can, and the hat. Then point and say", "Find the stickers. Stick the desk, the dog, the door and the duck. Then point and say", etc.

4. How to teach English to grade 1 and grade 2 children

4.1. Characteristics of grade 1 and grade 2 children

Researchers and second/foreign language teaching methodologists (e.g. Lenneberg, 1967; Stern 1967; Rivers, 1970; Broughton et al, 1978; Brown, 2000; Moon, 2005; Cameron, 2011; Wendy, Scott and Ytreberg, 2011; Ur, 2012; Fernández and Cairns, 2015; Shin and Scrandall, 2015) have suggested numerous characteristics of young second/foreign language learners. The following observations seem to be relevant to Vietnamese learners of six- and seven-years old:

• They are physically active.

• They understand meaning from what they see and hear.

• They have relatively short concentration and attention span.

• They can understand direct interaction between people.

• They like to play, sing, and mime.

• They understand situations faster than they understand the language being used.

• They are less aware of language forms and the possibility of making mistakes in these forms.

• They use and accept language more unconsciously.

• They can mimic very well.

• They do not have difficulty in pronouncing foreign language sounds, and in imitating rhythmic and intonation patterns.

• They are ready to form small groups to do an activity together.

4.2. Factors affecting grade 1 and grade 2 children's learning of English

The following factors may be said to affect Vietnamese six- and seven-years old children's learning English:

• They are enthusiastic and lively while doing things, but can be very likely to lose interest.

• English is an optional subject.

• The amount of experience of English given to them is extremely limited: 2 periods per week, 70 periods per year.

• English is a foreign language; the language of their daily communication is Vietnamese.

• The number of pupils per class is very large, usually over 40 per class; in some schools and localities, the number of pupils per class is even larger: over 50 or 60.

• When children start to learn English they are already fluent in oral Vietnamese: they already can use in their speech the Vietnamese sounds which we call phonemes such as /i/, /a/, /n/, /m/, etc.; they already can combine those phonemes to form what we call words to refer to things/objects which we call nouns such as *quả táo* (apple), *cái bàn* (table), *xe ca* (car), *túi xách* (bag), qualities which we call adjectives such as *to* (big), *cao* (tall), *lớn* (rộng), *vuông* (square), *tròn* (round), *dài* (long), *ngắn* (short), đẹp (beautiful), *xấu* (ugly), actions and states which we call verbs such as đi (go), *chạy* (run/move), *hát* (sing), *múa* (dance), etc.; they already can combine objects, qualities, actions into longer chunks which we call sentences such as *Có một chiếc xe to đang chạy trên đường* (There is a big truck moving on the road); they already can use these longer chunks to state, to negate, to ask, to answer, and even to argue with others in appropriate contexts of use. In short, when pupils start learning English, they can express orally in Vietnamese what they are doing without attending a formal class. When they go to school, they learn formally what has been referred to as "literacy": they learn how to read and write by first relating the spelling forms (which they are taught formally) to the sound forms (which they have already acquired informally) of Vietnamese; and then they learn how to write and read the spelling-sound forms on the page to produce and comprehend meaning. In contrast, when pupils start to learn English, they know nothing about the language both in its spoken and written forms; and they have to learn all things concerning these forms at the same time.

• Unlike countries of Kachru's (1985) "Outer Circle" where the learning of English takes place in both "natural environment" (one in which students can learn and acquire English both in and outside classroom) and "artificial environment" (one in which students learn English in classroom), the children's learning of English in Vietnam takes place in "artificial environment" which is often a very acquisition-poor environment.

• Children in urban and affluent areas often receive more attention and encouragement to learn English from parents than those in rural, remote and isolated areas.

• Teachers are recruited on a part-time basis; most of them are untrained in classroom methods and techniques of teaching English as a foreign language to younger learners.

4.4. Some suggested classroom methods of teaching English to grade 1 and grade 2 children

The world is replete with approaches to and methods of "language teaching" (one of the most slippery terms for the practical EFL teacher). If one attempts a quick search into books and reference materials in what is commonly referred to as "language teaching libraries" published in English, one can find numerous of them such as grammar-translation method, audio-lingual method, direct method, reading method, total physical response, silent way, suggestopedia, the lexical approach, competency-based language teaching, task-based language teaching (TBLT), and many others. How far it is possible to apply these methods and approaches to a specific foreign language teaching context, however, seems to be under-researched. The extent to which the classroom teacher can use these methods and approaches in his or her actual classroom teaching is, therefore, doubtful. The label most widely accepted in the "language teaching" world of today is "communicative language teaching" (CLT). CLT, for many scholars, is "an *approach*, not a method … . It is a "broadly based theoretical position about the nature of language and language teaching and learning" (Brown, 2000: 266). It encompasses a diverse set of principles reflecting a communicative view of language and language learning" (Richards and Rodgers, 2003: 172; see also Hammer, 2005: 84-5). And like all other language teaching methods and approaches, CLT is a big word and is too general to be applied to a specific classroom context (like the foreign language classroom in Vietnam) which uses a specific set of textbooks (like the VEPH-Macmillan set of textbooks for grade 1 and grade 2) to fulfil the goal(s) of a specific curriculum (like the *General School Education Introductory English Curriculum for Grade 1 and Grade* 2). This suggests that our method of teaching a particular course/programme should be specific. It should take into consideration all the specific features such as the children's linguistic, psychological, and physiological characteristics; their economic, cultural, and environmental conditions; the goal of *the Curriculum*, the content of the textbooks, the use the children will make of their ability in English once they have acquired it, the amount of English knowledge and the level of English skills they are required to achieve, the length of the course and the teaching hours in the course, the length of a period/lesson, the qualification of the teachers and their institutional and professional constraints, the tests or examinations imposed, and many others. With these specificities mind, we now move on to outline some classroom methods of teaching English to Vietnamese grade 1 and grade 2 children. The outline is organized into two headings: teaching methods and teaching procedure.

4.4.1. Teaching methods

Children "love to imitate and mime; they are uninhibited by acting out roles, and they enjoy repetition because it gives them a sense of assurance and achievement. They are ready to accept the fact that other people do and say things differently without worrying about the reasons for the differences. They love an active situation where they can express themselves vocally and physically. They enjoy singing and playing, and doing as the teacher does" (Rivers, 1970: 362). This being so, a multiple method should be employed to teach English to grade 1 and grade 2 children, which can be generally captured in the following four terms: **oral method**, **activity method, visual method** and **ostensive method**.

1. The oral method (somewhat similar to the traditional aural-oral or audio-lingual method) is the first to be recommended here because language is primarily a spoken activity and letters and prints are abstract symbols representing it. There seems to be some transfer from learning the spoken word to learning reading and writing, but very little indeed the other way. The well-mastered, correctly learned spoken language will help both subsequent reading and writing (Rivers, 1970; see also O'Byrne, 1973). By the oral method is meant that listening and speaking are taught first – listening, then speaking. Reading and writing will come in later. The linear order of listening – speaking – reading – writing, however, does not mean that these skills are taught separately one after another as suggested in a number language teaching methodology textbooks (e.g. Rivers, 1970; Richards and Rodgers, 2003; Harmer, 2005; Ur, 2012). Neither does it mean that the teaching of reading and writing in English should take place until a fluent oral foundation has been established as in the case of teaching the mother tongue or the first language. Rather it means that the four language skills of listening, speaking, reading, and writing should be taught either simultaneously in one and the same lesson, or in consecutive sequence in which listening and speaking may come in in the first lesson, and reading and writing in the next. It has become common that most of our children come to learn English with the expectation of speaking it and this expectation is intensely interested by their parents. If there were no other reasons but this, some effort should be made to give children successful oral start as a motivating factor. But surely our children cannot suddenly speak English without being given some oral language input. For them to be able to speak English, listening is the first skill that should be taught; and the first and normal step is that children receive the sounds – the main source of the language – through their ears (listen), and unlike the way they acquire the sounds of their mother tongue, after they hear the sounds of a foreign language they are required to repeat (speak) them.

The oral method can be used to teach the "Listen and repeat" activity effectively which is designed across the sixteen learning units in *English 1 – Student's Book*. It can also be effectively employed to teach chants, songs, and games. In *English 1 – Student's Book*, chants and songs are presented in lines. In teaching how to sing a chant or a song, the teacher first sings the chant or the song (word after word, then line after line) for the pupils to hear. Then she sings again and lets the pupils repeat the chant or the song in the same way as she did. Then she lets the pupils listen to the chant or the song either from her own performance or from the audio recording and say and repeat the chant or sing the song along in chorus until they can perform the task independently. Whatever criticism one may make about the drawback of the oral method, it still exists in the current FL classroom, and is still popular among EFL teachers and students all over the world. No one can speak another language without being taught in the oral method of one form or another. It is especially suitable for teaching young children as it encourages listening and repetition – activities that lead to oral fluency in English.

2. Children are physically active; "They love an active situation where they can express themselves vocally and physically" (Rivers, 1970: 362; see also Shin and Scrandall, 2015: 25). Here **the activity method** enters to be recommended. By the activity method is meant that language is best taught through activities, and verbal language (language contents and language skills) should be taught along with non-verbal language (body movement, facial expressions, gestures, mimes and actions). In *English 1 – Student's Book,* there are many activities for the teacher to employ the activity method effectively to teach his or her pupils: "Listen and repeat", "Point and say", "Listen and chant", "Listen and tick", "Trace", "Trace and then write", "draw and colour", "Let's talk", "Listen and tick" "Listen and match", "Find and circle", "Let's play (games such as "Simon says", "A happy circle", "Picture dominoes")". These activities are impressively presented and well woven throughout the book. They are fun, enjoyable, and natural to young children. Through these "learning and/by doing" activities, children can absorb and learn the language effectively.

Children are ready to form small groups to do an activity together. The readiness with which they form groups and participate in team activities is a quality which lends itself to the English lesson. Not only does group work give children more chance to talk with each other but it harnesses the purposeful and instrumental use of English (Broughton et al., 1978: 170). The activity method is in line with these characteristics of children: it matches language processing with physical action (Asher, 1977; Harmer, 2005); it makes language learning more effective when it is fun (Larsen-Freeman, 2003);

and, in particular, it sees language learning as an inter-subjective undertaking, focusing on the co-participation of the teacher and his or her pupils, and of one pupil with other(s) of his or her peers (Breen and Candlin, 1980).

3. "Lack of aural stimulus is relatively easy to tolerate; even young learners will work for a while in silence without searching for something to listen to. This, however, is not true of the visual, which is a very dominant channel of input: so much so, that if young learners are not supplied with something to look at that is relevant to the learning task in hand they will find and probably be distracted by something that is not" (Ur, 2012: 289). This suggests that **visual method** should form part of the total range of methods of teaching English to young children. By the visual method is meant that the teacher employs visual materials to teach language. Visual materials are of various types. The most obvious type for children is the picture; and the more clearly visible, striking and colourful the better. The teacher is advised to use the pictures in professionally drawn pictures or paragraphs: those are in the textbook, or coloured posters, or pictures cut from magazines. For those classes that are computer-assisted, the teacher can display the pictures he or she has selected on the screen and help the pupils do the activity/task.

It is obvious that classroom conditions impose a strict limit on living language situations, but the good teacher can seek to devise true-to-life situations for her pupils to experience authentic language in a contextualized way. Ways of devising situations may vary from the presentation of pictures, models to acting out a dialogue or a game. If, for instance, the teacher wants to teach school things, she may devise a school situation by making a large picture of a school in which *pens*, *books*, *notebooks*, *erasers*, *school bags* are presented. Or if the teacher wants to teach some foods and drinks, *fish*, *chips*, *milk*, and *chicken* can be selected to be taught by means of a picture of a "fish and chip" shop (see Unit 5, *English 1 - Student's Book*, Page 23).

4. It is generally agreed that in teaching a foreign language, one teaches meaning, not just form. The question of meaning is very complicated, and to teach young children a meaning in a language other than their own mother tongue may well be difficult for the foreign language teacher. To solve the problem, the teacher is advised to use a number of different strategies. In the early stage of English learning, she may say or explain in Vietnamese, or, when the language is short and simple, she may say in English first and then in Vietnamese. These teaching strategies may be faced with some impractical "theoretical objections", but in actual practice, they are quite effective considering the difficulty our young children of a foreign language meet when they hear the teacher says rubrics in English for the first time such as "Find the words.", "Write/Trace the letters.", or "Listen and tick", or the long and complex explanation of how a game is played (which may be even beyond the teacher's English language proficiency) such as the game "Pass the ball" (Lesson 1, Unit 6, *English 1 – Student's Book*). Which language (English or Vietnamese) should the teacher use to explain the rules of the game to her pupils when in Vietnamese the explanation would be roughly like this: "Bây giờ chúng ta cùng nhau chơi một trò chơi nhé. Tên trò chơi là "Pass the ball". Cô có luật chơi như sau: Cô có một quả bóng. Khi cô bật nhạc, cô sẽ chuyền quả bóng cho một bạn. Bạn đó phải chuyền quả bóng tiếp cho bạn thứ hai. Bạn thứ hai phải chuyền quả bóng tiếp cho bạn thứ ba. Cứ thế đến khi nhạc dừng lại, bạn nào cầm bóng không chuyền cho bạn khác khi nhạc vẫn đang bật phải nói to một từ trong bốn từ chúng ta vừa học: *bell*, *pen*, *pencil*, hoặc *red*. Các con rõ chưa nào?" (Now let us play a game. The game is "Pass the ball". The rules of playing this game is as follows: I have a ball. When I turn on the music, I will pass the ball to one of you. The one who receives the ball from me will have to pass it to another friend (second friend). Then he or she will pass the ball to another friend (third friend). And so you will continue passing the ball until the music stops. Whoever fails to pass the ball when the music is on must say aloud one of the four words we have just learned: *bell*, *pen*, *pencil*, or *red*. Are your clear?).

Alternatively, in order to teach meaning, the teacher may use what is commonly referred to as **the ostensive method** (which shares some features with the visual method). By the ostensive method is meant that the teacher teaches meaning through perception. She may point to and name people, (toy) animals, or objects such as Bill, Ba, Sue, Phil, car, cat, cup, dog, etc.; or she may make a simple contrast between people, animals, or objects (e.g. *short* girls v. *tall* girls, *big* cars v. *small* cars, *brown* dogs v. *black* dogs, etc.); or she can teach English prepositions by actually moving an object from "in" to "on" or "over" some other objects (e.g. *The dog is running* in *the room*. *The car is moving* on *the road*. *The man is jumping* over *the fence*).

Ostensive tools such as gestures and actions (those that are nowadays generally subsumed under the umbrella term of TPR (Total Physical Response) actions (Asher, 1977) can also be used to teach meaning. In teaching pupils how to sing a song, for example, the teacher announces what she is doing at the moment she is doing it (she sings the song and performs actions and gestures along). Pupils listen to the teacher's singing and observe her accompanying actions and gestures. In due course, they will take over the actions and gestures and at the same time will use the appropriate language. The procedure can be repeated until the children can sing the song independently. In this way, the teacher can put the language which is to be learned into authentic context of use so that the pupils may "live the language" (Wallwork, 1972: 153): they can understand meaning through experiencing the situation and the language that occurs in the situation (cf. Halliday et. al., 1964). Apart from actions and gestures, the teacher can use flash cards, pictures, and various objects and realia to make the classroom come alive and resemble more closely to the real world outside.

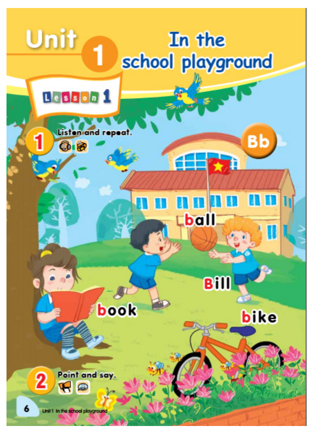
Proper use of the oral method, activity method, visual method, and ostensive method can free the teacher to be more the observer and facilitator in the learning process and free the pupils to learn to use the language and learn to love the potentially fabulous experience of being able to communicate in another language with other people around – a vivid example of the "learner-centred approach" in foreign language teaching (cf. Nunan, 1991; Tudor, 1996).

4.4.2. Teaching procedure

The specific classroom methods outlined above cannot be effective without the teacher's exploitation of an appropriate teaching procedure. To date, there are a number of teaching procedures available in the second/foreign language teaching literature. Apart from the classic tripartite procedure of **P-P-P** which stands for **P**resentation, **P**ractice, and **P**roduction, there are at least four other alternatives: (1) the **A-R-C** procedure (Scrivener, 1994) which stands for **A**uthentic use, **R**estricted use, and **C**larification and focus, (2) the **O-H-E** procedure (Lewis, 2002 cited in Harmer, 2005) which stands for **O**bserve, **H**ypothesize, and **E**xperiment,

(3) the **I-I-I** procedure (McCarthy and Carter, 1995) which stands for **I**llustration, **I**nteraction, and **I**nduction, and (4) the **E-S-A** procedure (Harmer, 2005, 2017) which stands for **E**ngage, **S**tudy, and **A**ctivate (for more detail of the **P-P-P** procedure and its alternatives, see Harmer, 2005: 80-4). A close examination of these teaching procedures reveal that there is nothing wrong with any one of them; the only problem with them may be that they are too general to be applied to the teaching of an EFL lesson which consists of several separate activities/tasks each of which is intended to achieve a particular communicative goal. The procedure we are suggesting here is more practical and more detailed, aiming to help the teacher how to deliver a specific activity/task effectively: the **G-I-P-O** procedure which stands for **G**oal, **I**nput, **P**rocedure, and **O**utcome. Details of this procedure can be elaborated as follows:

The teaching of an activity begins with the teacher's setting a **G**oal for the activity. After the goal has been defined, the teacher moves on to the second step – specifying language **I**nput (sounds, letters, words, including audio and visual materials) needed for teaching and learning the activity. Once the language input has been specified, the teacher moves on to the third step – the **P**rocedure – where the teacher actually handles the teaching point of the activity which consists of a number of sub-steps (how many steps needed depends on the nature of a particular activity). The activity ends with the **O**utcome step where the teacher checks her pupils' achievement against the goal that has been set. If the goal is achieved, the teacher can move on to the next activity. Below is an example of how an activity can be taught, using the **G-I-P-O** procedure. The activity is **Listen and repeat** taken from Lesson 1, Unit 1, *English 1 – Student's Book* (Page 6).

<table>
<tr><td>Goal:</td><td colspan="2">Pupils will be able to pronounce the sound of the letter B/b in isolation and in the words ball, Bill, book, and bike correctly.</td></tr>
<tr><td>Input:</td><td colspan="2">The teacher prepares input (both linguistic and audio and visual materials or realia) for teaching the activity: She may use the context picture (on Page 6), or she may enlarge the picture, make flashcards which contain both the upper case letter B and the lower case letter b, highlight them (either in bold or in colour) in the words ball, Bill, book, and bike, etc.</td></tr>
<tr><td>Procedure:</td><td>The teacher can break the procedure down into the following steps:
Step 1: the teacher has pupils look at the picture (saying Look at the picture, please!). Then she draws pupils' attention to the two modes of letter B/b, the word next to the ball and other words next to the things, and the colour of the letter B/b.
Step 2: the teacher asks pupils to point to the letter B/b. Then she plays the recording for them to hear and repeat the sound of the letter B/b (saying Listen and repeat, please!). (She can repeat the procedure, if necessary).</td><td>
</td></tr>
</table>

	Step 3: the teacher then asks pupils to point to the ball and then to the word *ball* (saying *Point to the ball and the word "ball" please*). Then she asks them to listen and repeat the word, (saying *Listen and repeat the word "ball", please!*) Step 4: the teacher repeats the same procedure with the words *Bill, book*, and *bike*. While the pupils practise, the teacher can give further support to them, especially to those pupils who find it difficult to do the task. Step 5: the teacher plays the recording again and asks the pupils to listen, point to the letter *B/b* and the word *Bill*, *book* and *bike* and repeat until they can perform the task effectively. Step 6: the teacher calls some pupils to pronounce the sound /b/ of the letter *B/b* in isolation and in the words *ball, Bill, book* and *bike* in front of the class (praising *Well done!* when the pupils perform the task well).
Outcome:	When the pupils can pronounce correctly the sound of the letter *B/b* in isolation and in the words *Bill, ball, bike* and *book*, the teacher can move on to the next activity (saying *Well done! Now we move on to the next activity*.).

5. Conclusion

This paper has been concerned with the new *General School Education Introductory English Curriculum for Grade 1 and Grade 2* promulgated by MoET in 2018. We claimed that *the Curriculum* is a new innovation in foreign language general school education in Vietnam because, as far as we know, although it is an optional curriculum, it is issued at the national level; no other foreign language curricula of similar kind have ever been issued at national level like this one in Vietnam. To highlight the innovation, we have explored some of the most prominent aspects of *the Curriculum*: the characteristics of *the Curriculum*, the principles for designing *the Curriculum*, the goal and objectives of *the Curriculum*, and the content of *the Curriculum*. Then we moved on to present in some detail the new set of English textbooks for Vietnamese grade 1 and grade 2 children developed by VEPH in collaboration with Springer Nature Macmillan Education, looking specifically at *Tiếng Anh 1* (English 1) which includes Student's Book, Teacher's Book, and Workbook. It can be seen that the general approach to developing the textbook is for communication. It is designed around sixteen learning units, four fun time sections, and four review units. Each learning unit is the centre around which language components (phonics/sounds, vocabulary, structures) and communication-based activities such as "Look, listen and repeat", "Listen and chant", "Let's chant", "Let's sing", etc., are organized. These language components and activities are integrated into the fabric of the lessons to facilitate pupils' development of listening, speaking, reading, and writing in English.

Realizing that without a classroom-specific teaching methodology and a proper teaching procedure, the teaching of English as a foreign language to children would cease to be ineffective, we have suggested four specific methods and a procedure for teaching an activity which we thought would stand benefit to the primary teacher of English: the oral method, the activity method, the visual method, and the ostensive method, and the G-I-P-O procedure. It should be noted at this point that although these teaching methods and the teaching procedure are beneficial to the teacher, they are only an artificial breaking up of what is actually an organic process from the beginning to the end of the lesson or activity. When it comes to actual teaching, much more should be done on the part of the teacher to make the lesson or the activity more enjoyable and effective.

It may be argued that the methods we have suggested favour imitation – pupils learn by imitating the teacher, using language. This is true since the actual classroom is only a small piece of the world in which we expect our children to use English, artificial means must be used to transform it into a variety of other pieces, and the obvious means for performing this transformation are "listening and repeating bits of the language", "listening, repeating and singing songs", and "listening, repeating, and singing chants". These have always been useful educational or teaching devices for children at the early stage of learning. By creating a variety of situations in the classroom – in part by letting pupils listen to and repeat the English sounds and then by letting them relabel or rename things/objects and people and relate them to the symbols (letters and words), but also in part by simply helping them act out small dialogues, sing simple songs and simple chants, and play simple language games as a class or in groups (supplemented by realia if desired) – the teacher can expand the classroom indefinitely, provide natural contexts for the language being used, and prepare her pupils for real communication.

The prominent features of *the Curriculum* and the textbooks we have presented, and the classroom methods of teaching we have suggested in this paper, no matter how important they are, are only three factors contributing to the success of *the Curriculum*. For *the Curriculum* to be effectively implemented, the following problems derived from the nature of *the Curriculum* itself should be solved:

First, because English is an optional subject, many primary schools do not find the need to organize the programme.

Secondly, because English is an optional subject, it is not tuition-free; many people tend to think that it is only suitable for schools in urban and affluent areas.

Thirdly, because *the Curriculum* is a new one, there will certainly be a shortage of teachers.

Fourthly, because English is an optional subject, it will be very difficult for schools to recruit qualified teachers.

And finally, because English is an optional subject and there is very little class time allocated to it, schools are not allowed to recruit teachers on a permanent basis. The result is that the teachers recruited will feel unassured of their work. This, combined with the poor salary they receive, will result in the ineffectiveness of *the Curriculum*, opening the success of the innovation to questions.

References

Vietnamese

Bộ Giáo dục và Đào tạo [MOET] (2010a). *Chương trình tiếng Anh thí điểm tiểu học* (Pilot English Curriculum for Primary Schools in Vietnam) ban hành theo Quyết định số 3321/QĐ-BGDĐT ngày 12 tháng 8 năm 2010 của Bộ trưởng Bộ Giáo dục và Đào tạo.

Bộ Giáo dục và Đào tạo [MoET] (2010b). *Quyết định số 4674/QĐ-BGDĐT ngày 5 tháng 10 năm 2010 của Bộ trưởng Bộ Giáo dục và Đào tạo về việc triển khai Chương trình thí điểm tiếng Anh tiểu học* (Decision on the Implementation of the Pilot English Curriculum for Primary Schools in Vietnam).

Bộ Giáo dục và Đào tạo [MoET] (2012a). *Chương trình giáo dục phổ thông môn tiếng Anh thí điểm cấp trung học cơ sở* (Pilot English Curriculum for Lower Secondary Schools in Vietnam) ban hành theo Quyết định số 01/QĐ-BGDĐT ngày 03 tháng 01 năm 2012 của Bộ trưởng Bộ Giáo dục và Đào tạo.

Bộ Giáo dục và Đào tạo [MoET] (2012b). *Chương trình giáo dục phổ thông môn tiếng Anh thí điểm cấp trung học phổ thông* (Pilot English Curriculum for Upper Secondary Schools in Vietnam) ban hành theo Quyết định số 5290/QĐ-BGDĐT ngày 23 tháng 11 năm 2012 của Bộ trưởng Bộ Giáo dục và Đào tạo.

Bộ Giáo dục và Đào tạo [MoET] (2012c). *Quyết định số 3456/QĐ-BGDĐT ngày 05 tháng 9 năm 2012 về việc triển khai Chương trình giáo dục phổ thông môn tiếng Anh thí điểm cấp trung học cơ sở* (Decision on the Implementation of the Pilot English Curriculum for Lower Secondary Schools in Vietnam).

Bộ Giáo dục và Đào tạo [MoET] (2013). *Quyết định số 3702/QĐ-BGDĐT ngày 10 tháng 9 năm 2013 về việc triển khai Chương trình giáo dục phổ thông môn tiếng Anh thí điểm cấp trung học phổ thông* (Decision on the Implementation of the Pilot English Curriculum for Upper Secondary Schools in Vietnam).

Bộ Giáo dục và Đào tạo [MoET] (2018a). *Chương trình giáo dục phổ thông: Chương trình tổng thể* (General School Education Curriculum) ban hành kèm theo Thông tư số 32/2018/TT-BGĐT ngày 26 tháng 12 năm 2018 của Bộ trưởng Bộ Giáo dục và Đào tạo.

Bộ Giáo dục và Đào tạo [MoET] (2018b). *Chương trình giáo dục phổ thông môn tiếng Anh* (General School Education English Curriculum) ban hành kèm theo Thông tư số 32/2018/TT-BGDĐT ngày 26 tháng 12 của Bộ trưởng Giáo dục và Đào tạo.

Bộ Giáo dục và Đào tạo [MoET] (2018c). *Chương trình giáo dục phổ thông làm quen tiếng Anh lớp 1 và lớp 2* (General School Education Introductory English Curriculum for Grade 1 and Grade 2) ban hành kèm theo Thông tư số 32/2018/TT-BGDĐT ngày 26 tháng 12 năm 2018 của Bộ trưởng Bộ Giáo dục và Đào tạo.

Hoàng Văn Vân (2016). Vai trò và vị thế của tiếng Anh trong thời kì hội nhập và một số hàm ý cho việc hoạch định chính sách ngoại ngữ ở Việt Nam (The role and status of English in the period of international integration and some implications for foreign language policy making in Vietnam). *VNU Journal of Science: Foreign Studies*, *46*, 12-27.

Hoàng Văn Vân (2019). Giáo dục ngoại ngữ ở bậc phổ thông trên thế giới: Tình hình hiện tại và những vấn đề. *VNU Journal of Foreign Studies*, *35*(5), 1-40.

Hoàng Văn Vân et al., (2019a). *Tiếng Anh 1 – Sách học sinh* (English 1 – Student's Book). Hà Nội: Nhà xuất bản Giáo dục Việt Nam.

Hoàng Văn Vân et al., (2019b). *Tiếng Anh 1 – Sách giáo viên* (English 1 – Teacher's Book). Hà Nội: Nhà xuất bản Giáo dục Việt Nam.

Hoàng Văn Vân et al., (2019c). *Tiếng Anh 1 – Sách bài tập* (English 1 – Workbook). Hà Nội: Nhà xuất bản Giáo dục Việt Nam.

Nhà xuất bản Giáo dục Việt Nam (Vietnam National Publishing House). sachmem.vn.

English

Asher, J. (1977). *Learning Another Language Through Actions: The Complete Teacher's Guide Book*. Los Gatos, Calif.: Sky Oaks Production.

Breen, M.P. and Candlin, C.N. (1980). The Essentials of a Communicative Curriculum in Language Teaching. *Applied Linguistics*, *1*(2), 89-112.

Broughton, G. et al. (1978). *Teaching English as a Foreign Language*. London, Henley and Boston: Routledge & Kegan Paul.

Brown, H.D. (2000). *Principles of Language Learning and Teaching*. New York: Pearson.

Cameron, L. (2011). *Teaching Languages to Young Learners*. Cambridge: Cambridge University Press.

Cheshire, J. (1996). Introduction: Sociolinguistics and English around the World. (In) *English around the World: Sociolinguistic Perspectives*, pp. 1-12. Cheshire, J. (Ed.). Cambridge: Cambridge University Press.

Eurydice (2005). *Key Data on Teaching Languages at School in Europe*. 2005 edition. Brussels: Eurydice.

Eurydice (2017). *Eurydice Report: Key Data on Teaching Languages at School*. Retrieved 18 Feb. 2018 from https://www.csee-etuce.org/en/news/archive/2038-eurydice-report-key-data-on-teaching-languages-atschool-in-europe

Fernández, E.M. and Cairns, H.S. (2015). *Fundamentals of Psycholinguistics*. Singapore: Wiley-Blackwell.

Halliday, M.A.K. (2017). Written Language, Standard Language, Global Language. (In) *Halliday in the 21st Century*, pp. 87-103. Webster, J.J. (Ed.). London: Bloomsbury.

Halliday, M.A.K., McIntosh, A., and Strevens, P.D. (1964). *The Linguistic Sciences and Language Teaching*. London: Longmans.

Harmer, J. (2005). *The Practice of English Language Teaching*. Third Edition. Malaysia: Longman.

Harmer, J. (2017). *How to Teach English*. New Edition. Thirteenth Impression. England: Pearson Education Limited.

Hoang, V. V. (2012). The Role of Textbooks in the Implementation of the National Project "Teaching and Learning Foreign Languages in the National Education System Period 2008-2020". *VNU Journal of Science: Foreign Studies,* 30, 75-89.

Hoang, V.V. (2018). MoET's Three Pilot English Language Communicational Curricula for Schools in Vietnam: Rationale, Design and Implementation. *VNU Journal of Foreign Studies*, *34*(2), 1-25.

Kachru, B.B. (1985). Standards, Codification and Sociolinguistic Realism: The English Language in the Outer Circle. (In) *English in the World: Teaching and Learning the Language and Literatures*, pp. 11-30. Quirk, R. and Widdowson, H.G. (Eds.). Cambridge: Cambridge University Press and the British Council.

Larsen-Freeman, D. (2003). *Techniques and Principles in Language Teaching*. Second Edition. Seventeenth Impression. Oxford; Oxford University Press.

Lenneberg, E.H. (1967). *Biological Foundations of Language*. New York: Wiley.

Lewis, M. (2002). *The Lexical Approach: The State of ELT and a Way Forward* (Language Teaching Publications). First Edition. Heinly ELT.

Littlewood, W. (1999). *Communicative Language Teaching*. Cambridge: Cambridge University Press.

McCarthy, M. and Carter, R. (1995). Spoken Grammar: What is it and How can We Teach it? *ELT Journal* 49/3.

Moon, J. (2005). *Children Learning English. A Guidebook for English Language Teachers*. Thailand: Macmillan Education.

Nunan, D. (1991). *The Learner-Centred Curriculum*. Fourth Printing. Cambridge: Cambridge University Press.

Nunan, D. (2003). The Impact of English as a Global Language on Educational Policies and Practices in the Asia-Pacific Regions. *TESOL QUARTERLY*, *37*(4) 4.

O'Byrne, V. de R. (1973). Curriculum Planning for Modern Languages. (In) *The Teacher's Role in Curriculum Design*, pp. 145-70. Hughes, P. (Ed.). Sydney: Angus and Roberton Publishers.

O'Neill, R. (1995). Why Use Textbooks? (In) *Current Change in English Language Teaching*, pp. 148-56. Oxford: Oxford University Press.

Pew Research Center (2015). *Learning a Foreign*

Language a 'must' in Europe, not so in America. Retrieved 22 Feb. 2019 from http://www.pewresearch.org/fact-tank/2015/07/13/learning-a-foreign-languagea-must-in-europe-not-so-in-america/

Richards, J.C. (2001). *Curriculum Development in Language Teaching*. Cambridge: Cambridge University Press.

Richards, J.C. and Rodgers, T.S. (2003). *Approaches and Methods in Language Teaching*. 6th Printing. Cambridge: Cambridge University Press.

Rixon, S. and Papp, S. (2018). The Educational Contexts of the Teaching of English to Young Learners and the Roles of Assessment. (In) *Studies in Language Testing 74*, pp. 18-69. Cambridge: Cambridge University Press.

Rivers, W. M. (1970). *Teaching Foreign Language Skills. Third Impression*. Chicago and London: The University of Chicago Press.

Scott, W.A. and Ytreberg, L.H. (2011). *Teaching English to Children*. Malaysia: Longman.

Scrivener, J. (1994). *PPP and after*. The Teacher Trainer 8/1.

Scrivener, J. (2009). *Learning Teaching: A Guide Book for English Language Teachers*. Second Edition. Oxford: MacMillan.

Shin, J.K. and Scrandall, J.J. (2015). *Teaching Young Learners English: From Theory to Practice*. Singapore: Heinle.

Stern, H.H. (1967). *Foreign Languages in Primary Education: The Teaching of Foreign or Second Languages to Young Learners*. Oxford: Oxford University Press.

Tudor, I. (1996). *Learner-centredness as Language Education*. Cambridge: Cambridge University Press.

Ur, P. (2012). *A Course in Language Teaching*. Cambridge: Cambridge University Press.

Wallwork, J.F. (1972). *Language and Linguistics*. London: Heinemann.

Wilkins, D.A. (1976). *Notional Syllabuses*. London: Oxford University Press.

"CHƯƠNG TRÌNH GIÁO DỤC PHỔ THÔNG LÀM QUEN TIẾNG ANH LỚP 1 VÀ LỚP 2": ĐIỂM CÁCH TÂN TRONG GIÁO DỤC NGOẠI NGỮ Ở TRƯỜNG PHỔ THÔNG VIỆT NAM

Hoàng Văn Vân

Trung tâm Nghiên cứu giáo dục ngoại ngữ, ngôn ngữ và quốc tế học,
Trường Đại học Ngoại ngữ, ĐHQGHN,
Phạm Văn Đồng, Cầu Giấy, Hà Nội, Việt Nam

Tóm tắt: Bài viết này liên quan đến một điểm cách tân quan trọng trong giáo dục ngoại ngữ ở bậc phổ thông Việt Nam đã và đang thu hút sự quan tâm đáng kể của các nhà giáo dục ngoại ngữ, các giáo viên ngoại ngữ, và công chúng – *Chương trình giáo dục phổ thông làm quen tiếng Anh lớp 1 và lớp 2* được Bộ Giáo dục và Đào tạo ban hành năm 2018. Để làm nổi bật điểm cách tân quan trọng này, bài viết dự định giải quyết ba vấn đề chính sau đây: (1) bài viết sẽ nghiên cứu và làm rõ một số khía cạnh quan trọng nhất của *Chương trình giáo dục phổ thông làm quen tiếng Anh lớp 1 và lớp 2*; (2) bài viết sẽ trình bày và thảo luận chi tiết các nội dung của *Tiếng Anh 1* – một bộ sách giáo khoa tiếng Anh mới được Nhà xuất bản Giáo dục Việt Nam phối hợp với Nhà xuất bản Macmillan Education thuộc tập đoàn Springer Nature biên soạn theo *Chương trình giáo dục phổ thông làm quen tiếng Anh lớp 1 và lớp 2*; và (3) bài viết sẽ sơ thảo và thảo luận một số phương pháp dạy học cụ thể trên lớp và trình bày một quy trình lên lớp mới để dạy một hoạt động / nhiệm vụ nhằm giúp giáo viên và học sinh dạy và học tiếng Anh có hiệu quả.

Từ khoá: Chương trình giáo dục phổ thông làm quen tiếng Anh lớp 1 và lớp 2, bộ sách giáo khoa *Tiếng Anh 1*, các phương pháp dạy học tiếng Anh trên lớp có hiệu quả, quy trình mới để dạy một hoạt động ngôn ngữ

HETEROGLOSSIA: ANOTHER SFG-BASED APPROACH TO TREATMENT OF WORD ORDER AS A MEANS FOR EXPRESSING MODALITY IN VIETNAMESE

Nguyen Van Hiep*

Vietnam Institute of Linguistics|
9 Kim Ma Thuong Street, Ba Dinh, Hanoi

Received 13 February 2020
Revised 26 April 2020; Accepted 18 July 2020

Abstract: The paper first introduces heteroglossia – a development of Systemic Functional Grammar (SFG) in Hallidayan Systemic Functional Linguistics (SFL) school before suggesting possible uses of heteroglossia in present-day studies of language, with particular focus on the role of word order in sentences of Vietnamese, a typical isolating language. The change of word order is considered a means for expressing modality, as shown in several interesting examples in Vietnamese, which proves that SFG and heteroglossia as its variant is an effective approach for exploring the role of word order in Vietnamese.

Keywords: isolating language, word order, modality, Systemic Functional Grammar (SFG), heteroglosia, grammaticalization.

Among the achievements of modern grammar, Systemic Functional Grammar (SFG) is a good grammar model. Since M.A.K Halliday (1985) built it up, SFG has evolved with various variants such as the evaluation theory and the heteroglossia approach. In this article, we first introduce heteroglossia as a development of SFG. Then we will discuss the role of word order as one of the most important grammatical means for making meaning in the Vietnamese language. Especially, we will focus on the role of word order as a means for expressing modal meanings within the framework of the heteroglossia approach.

1. Heteroglossia approach as a development variant of SFG

1.1. A synopsis of heteroglossia approach

The heteroglossia approach in modality studies originates from the interpersonal meaning as one of the three aspects of the sentence in SFG framework. This approach was proposed by White (2003, 2006) in two papers, which are "Beyond modality and hedging: A dialogic view of the language of intersubjective stance" (2003) and "Dialogistic positioning and interpersonal style - a framework for stylistic comparison (co-author with Motoki, 2006). White claims to have taken inspiration from the views of two Russian poetics researchers Bakhtin and Voloshinov on the dialogue of all kinds of discourse, whereby "verbal interaction is the basic reality of language. Dialogue . . . can also be understood in a broader sense, meaning not only direct, face-to-face, vocalised verbal communication between persons, but also verbal communication of any type whatsoever. A book, i.e., a verbal performance in print, is also an element of verbal communication. . . . [it] inevitably orients itself with respect to previous performances in the same sphere . . . Thus the printed verbal performance engages, as it were, in ideological colloquy of a large

* Tel.: 0904763131
Email: nvhseoul@gmail.com

scale: it responds to something, affirms something, anticipates possible responses and objections, seeks support, and so on" (Voloshinov, 1995, p.139).

Bakhtin similarly observes that all utterances exist

> … against a backdrop of other concrete utterances on the same theme, a background made up of contradictory opinions, points of view and value judgements … pregnant with responses and objections (1935 [1981], p.281])

The heteroglossia viewpoint is also influenced by Martin (Martin and White, 2005), who has the same semantic and rhetoric orientation when proposing the concept of "engagement" as a comprehensive category of linguistic resources to express interactive, inter-subjective perspectives in evaluation theory.

In short, just as SFG always associates sentence research in relation to discourse, the heteroglossia viewpoint assumes that the study of modality cannot be confined to the attitude and judgment of the speaker in relation to the content of propositions, as Lyons (1977) and Palmer (2001) conceived. Instead, White and Motoki declared, 'Thus while earlier treatments have tended to interpret modals and evidentials as signs of lack of commitment by the speaker to the truth value of the proposition, we are directed, rather, to attend to the intersubjective, dialogistic effects associated with such meanings' (White and Motoki, 2006).

So, from a heteroglossia standpoint, White and Motoki accepted a broad understanding of modality for analysing the linguistic resources of intersubjective positioning. They argued that linguistic means have long been thought to represent, *inter alia*, polarity, evidentiality, hedging, concession, intensification, authority, consequentiality, all of which can be grouped under the modality term. On the basis of discourse semantics, they all provide the means for speakers/writers to take a stance towards the various points-of-view or social positionings being referenced by the text and thereby to position themselves with respect to the other social subjects who hold those positions. It can be said that the heteroglossia approach is an interesting development of modality studies, which put modality in a dialogue perspective, and attach modality to the situations of the discourse.

Heteroglossic utterances are also distinguished by White and Motoki (2006) from monoglossic ones. A monoglossic utterance is a case in which the utterance does not show signs of acknowledging alternative views or there is no awareness of such viewpoints, expicit or potential in the dialogue. From the Bakhtinian perspective, such an utterance is "monoglossic" or "undialogized" assertion (Bakhtin, 1935 [1981], p.427). For example, in Vietnamese, the utterance

(1) Ngày mai nó đến.
tomorrow it comes
'Tomorrow he/she will come.'

is monoglossic, with categorical assertion, distinguished from utterance (2)

(2) Ngày mai thế nào nó cũng đến.
tomorrow no matter what it also come
'Tomorrow he/she will definitely come [no matter what happens].'

because of different perspectives on the possibility of his/her coming or not. On the contrary, a heteroglossic utterance, according to White (2003), is the one that expresses the argument with a different point of view or stance. The idiomatic expession "thế nào... cũng" (whatever/no matter what … also) in the utterance is an indicator of such an argument.

1.2. Two kinds of heteroglossic engagement: dialogistic contraction and expansion

To clarify the nature of modality, White and Motoki (2006) coined the term "heteroglossic engagement" and attributed all linguistic resources expressing heteroglossic engagement to two broad categories, namely, dialogic expansion and dialogic contraction. White also developed a set of terminology to clarify the nature of heteroglossia. The following presentation is the most general introduction to this terminology set with necessary interpretations and illustrations in Vietnamese.

1.2.1. Dialogic contraction

Statements containing *dialogic contraction* have indicators to prevent or narrow the space for alternative viewpoints, even though there can be several. For example, in Vietnamese, when advising someone to end a relationship with someone else, one may say,

(3)	*Đằng nào*	cậu	*cũng*	*phải* cắt quan hệ	với	con người	đó.
	anyway	you	also	must cut relation	with	person	that

'Anyway/all things considered, you have to terminate relationship with that person.'

By using "đằng nào... cũng phải", the speaker excludes arguments that the interlocutor can offer to reject the advice.

Dialogic contraction is represented by PROCLAMATION and the DISCLAMATION. As for PROCLAMATION, utterances contain indications that the speaker has individual 'investments' in the stated point of view, and is interested in raising that view as if to refute the opposing viewpoint. For example,

(4)	Tôi	*buộc*	*phải*	nói rằng	việc	đó	chẳng	ra	gì.
	I	force	must	say that	thing	that	not	out	what

'I am obliged to say that that is nonsense.'

the phrase "*buộc phải*" indicates that this is an affirmative, public and authoritative statement and the speaker intends to deny all opposing views. In example (5) below,

(5)	Ông ấy	đã	phản bác	*chuẩn*	*không*	*cần*	*chỉnh*	đề án	đó.
	he	PAST	refute	standard	no	need	adjust	project	that

'He refuted that project rightfully [without any need for adjustment].'

the phrase "*chuẩn không cần chỉnh*" shows the speaker's disagreement with the project.

As regards DISCLAMATION, the utterance contains indications of rebuttal, confrontation, or challenge to opposing views. For example,

(6)	*Tội gì*	mà	cô	cứ	cung phụng	tiền bạc	cho	anh ta	đánh bạc.
	sin what	CON	FEM	still	donate	money	give	him	gamble

'You don't have to trouble yourself by offering him money to gamble.'

(CON = connective; FEM = a form for addressing women)

the phrase "*Tội gì*" indicates that the speaker rejects any deontic basis that justifies the woman's offering of money for the man to gamble. In another instance,

(7)	*Sự thật*	*là*	tôi đã	không hề	biết	chuyện	này.
	truth	be	I PAST	no whatsoever	know	thing	this

'The truth is I do not know about it at all.'

the phrase "*Sự thật là* " indicates the speaker's strong opposition to a different view.

1.2.2. Dialogistic expansion

In the category of *Dialogistic expansion*, White and Motoki (2003) said that this is a case of utterances with indicators that different views are alternative and the difference between them is only in terms of the degree of epistemic modality. For example, in Vietnamese, when I say,

(8)	*Tôi*	*tin*	rằng	mọi	chuyện	sẽ	ổn.
	I	believe	that	all	thing	will	fine

'I believe that everything will be fine.'

the phrase "*Tôi tin*" indicates other possibilities, e.g. there may be someone who doesn't share my view, someone who thinks everything will be bad.

The category of *dialogistic expansion* comes in two types, which White (2003) calls ENTERTAIN and ATTRIBUTE. As for ENTERTAIN, the utterance contains indications that the speaker makes a conditional statement, which is only one of the possibilities. For example, in the utterance below,

(9)	*Có lẽ*	tôi	sẽ	chuyển	vào Sài Gòn.
	Maybe	I	will	move	in

'Maybe I will move to Saigon.'

the phrase "*Có lẽ*" indicates that my move to Saigon is only one among different possibilities (e.g., I may still stay in Hanoi).

Concerning ATTRIBUTE, the speaker says what he/she says from a certain point of view as one among different possibilities, and its authenticity depends on the evidence or credibility of the owner of that point of view. For example, in the following utterance:

(10)	Chính quyền	thành phố	*khẳng định*	tệ mại dâm	đã	chấm dứt.
	government	city	confirm	prostitution	PAST	end

'The city government confirms that prostitution has ended.'

the phrase *"khẳng định"* indicates that the situation that "prostitution has ended" is just a statement from the city government. The speaker points out one possibility, leaving space for other opinions (e.g., the opinion that prostitution is still ongoing, or has changed into more sophisticated forms).

2. An overview of the roles of word order in Vietnamese

As is known, language linearity means, in a way, that any changes of word order result in various syntactic, semantic and pragmatic changes, *inter alia,* in natural languages. Therefore, all languages use changes of word order as a means of expression. However, this method is differently applied across languages. Hereafter is a discussion of word order in Vietnamese.

In terms of typology, Vietnamese is an isolating, or analytic language. Like other languages of the same type, word order plays a crucial role in expressing meanings. While learning Vietnamese, foreign learners might be surprised when being asked to reorder the five words *sao (why), bảo (say), nó (it), không (not), đến (come)* in the senetence

(11) Sao bảo nó không đến.

'Why did you say that he would not come?

because numerous grammatically correct and acceptable sentences can be created from that original sentence, such as:

(11a) Sao nó bảo không đến?

'Why did he say he would not come?'

(11b) Nó bảo sao không đến.

'He said, "Why don't you come?

(11c) Không sao bảo nó đến.

'No problem, tell him to come.'

(11d) Nó đến bảo không sao.

'He came to say "no problem!"'

(11e) Không đến, nó bảo sao?

'Couldn't come. What did he say?'

(11f) Đến nó, sao không bảo?

'Why didn't you tell me you came to him?'

The important point is that, for SFG based on which the heteroglossia approach is developed, grammar is a system of choices for expressing meaning, and word order differences are also options for expressing meaning.

2.1. Word order, grammatical functions and representational meaning

According to traditional grammar, grammatical functions are primarily determined by a word's position in a sentence, namely subject, verb, object, complement, etc. Representational meaning is conveyed through semantic roles, i.e. the roles of words that create a state of affairs in a sentence. At the sentence level, the change of word order obviously leads to the change of their grammatical functions and ultimately the change in the sentence's representational meaning. For example,

(12) Tôi$_1$ đánh nó.
I hit it
'I hit him.'

(13) Nó đánh tôi$_2$.
'He hit me.'

'tôi$_1$' in (12) is the subject of the sentence, assuming the semantic role of the agent of the action denoted by the verb 'đánh' while 'tôi$_2$' in (13) is the object in the semantic role of the patient of the action.

Similarly, within a syntagm, or a phrase, a change in word order will effect a change in the syntagmatic functions of the elements and consequently the representational meaning of that syntagm or phrase. Cf. *con gà/gà con* (a chicken or a rooster or a hen/a chick), *hai vợ/ vợ hai* (two wives/ second wife). Furthermore, the change of word order also leads to changes in modality, tenses and aspects. For example, the word *được*'s meanings substantially vary in the following sentences:

(14) Nó **được** nhà, **được** vợ.

'He **has** a house, **has** a wife.'

(15) Anh ấy **được** đi chơi.

'He **has been** allowed to hang around.'

(16) Quả này ăn **được**, không chết.

'This fruit **can** be eaten, no poison.'

(17) Hôm qua, chị ấy mua **được** cái áo rất đẹp.

'Yesterday, she bought a nice shirt **already**.'

(18) Làm thế là **được**.

'That's **fine**.'

(19) Cô ấy **được** thầy khen.

'She **was given** nice compliments by her teacher.'

Although researchers might argue over the grammatical functions of the word *được* in the above examples, it is easily recognized that in (14) ***được (has)*** is a verb indicating possession while ***được (has been)*** in (15) is a passive form indicating allowance in terms of deontic modality; ***được (can)*** in (16) is a modal verb indicating possibility in terms of espistemic modality; ***được (already)*** in (17) is an adverb indicating completion; in (18), ***được (fine)*** is an assessment adjective; and finally ***được (was)*** plays a role of a passive form in (19).

2.2. Word order and topicalisation

The change of word order is also applied in the process of topicalization, which creates the topic (theme) of a sentence (Nguyễn Kim Thản, 1964; Nguyễn Văn Hiệp, 2009).

(20a) Ông ấy không hút thuốc.
He not smoke cigarettes.

'He does not smoke cigarettes.'

(20b) →Thuốc, ông ấy không hút Ø.

'Cigarettes, he does not smoke Ø.'

(21a) Bà ấy có hàng dãy nhà ở phố, hàng mẫu ruộng ở quê.
She has rows house live street acres fields live countryside

'She has rows of houses in the city, and acres of farms in the countryside.'

(21b) → Nhà, bà ấy có hàng dãy Ø ở phố; ruộng, bà ấy có hàng mẫu Ø ở quê.

'Houses, she has rows Ø in the city; farms, she has acres Ø in the countryside'.

3. Applying heteroglossia approach to word order in relation to modality in Vietnamese

Using word order change to transform an expression, resulting in changes in meaning, is the strategy used by all languages. What matters is the scales of application and the types of meaning created by the changes of word order, which plays varying roles in different language types. One of the types of meanings that are created by changing word order is that of a modality, conversational meaning when there are different points of view. Therefore, the heteroglossia approach can be applied to studying various types of modality meanings created by word order change. For example, in Vietnam, the heteroglossia approach has been applied to examining the system of final modal particles and the system of modality idioms (Nguyễn Văn Hiệp, 2018, 2019). However, within the scope of this article, we restrict ourselves to application of the heteroglossia approach to exploring the types of modality meanings conveyed when we change the word order in phrases and sentences in Vietnamese.

3.1. Change of word order in phrases

Applying the heteroglossia approach to Vietnamese enables us to explain the kinds of modality meanings brought about by the changes in word order. For example, swapping the adverb "lại" (again) and a verb in a verb phrase results in completely different constructions of "verb + lại" and "lại + verb", which can be serious challenges to learners of Vietnamese. The reason is *lại (again)* conveys different meanings when being put before or after a verb, as in

(22) Nó	**lại**	học.
It	again	learn

'He continues to learn **again**.'

(23) Nó học **lại**.

'He repeats [the grade].'

In (22), apart from describing a repeated action, the word ***lại*** also expresses the speaker's subjective judgement (a type of modality upon the speaker's view) together with an implication that the learning activity is unexpected and somewhat negative or worrisome/annoying to the speaker. This is the meaning of ***lại*** when occurring before a verb, as in *Nó **lại** hỏi mượn tiền* (*He asked for money **again***); *Nó **lại** đánh vợ* (*He hit his wife **again***); *Nó **lại** hút thuốc* (*He smokes **again***)… However, in (23), ***lại*** describes a repeated action without any explicit judgment nor implied annoyance on the part of the speaker. This is the meaning of *lại* when being put after a verb, like *Nó làm **lại** bài toán (She does her math exercises again); Cô ấy nói **lại** câu đã nói hôm qua (She repeats what she said yesterday)*.

By contrast, in cases like the following, ***lại*** conveys the speaker's subjective assessment and attitude to the actions or states of affairs in the sentence. For example:

(24) Giữa lúc gia cảnh túng quẫn vì mẹ ốm triền miên, anh con cả **lại** lấy vợ.

'The trouble his family was suffering from due to his mother's sudden sickness worsened with his first son's marriage.'

(25) Giữa lúc ông cần một trợ thủ để chống lại nạn ăn cắp vật liệu tại công trường, con chó **lại** lăn ra chết.

'When he was in need of help to combat against thefts at the construction site, his dog died.'

Another interesting case includes a construction in which an adjective precedes a noun referring to human body parts. That construction can be a subject-verb structure or a noun phrase, and when the word order is changed, with the adjective following the noun, the construction may be a mere adjective phrase with apparent symbolic, idiomatic or figurative meanings. For instance, *tay mát* vs. *mát tay* (lit. *hand [is] cool* vs. *cool hand*, which means people with dexterity and/or capacity to do things effectively); *mặt mát* vs. *mát mặt* (lit. *face [is] fresh* vs. *fresh face,* meaning satisfied, happy or proud); *mặt đẹp* vs. *đẹp mặt* (lit. *face [is] beautiful* vs. *beautiful face,* meaning proud); *mặt vàng* vs. *vàng mặt* (lit. *face [is] yellow* vs. *yellow face*, meaning scared or starving); *mặt xanh* vs. *xanh mặt* (lit. *face [is] green* vs. *green face*, meaning scared); *gan to* vs. *to gan* (lit. *liver big* vs. *big liver*, meaning brave, courageous); *bụng tốt* vs. *tốt bụng* (lit. *abdomen [is] good* vs. *good abdomen*, i.e. nice and kind [people]); *đầu to* vs. *to đầu* (lit. *head [is] big* vs. *big head*, i.e. grown-up; *mắt xanh* vs. *xanh mắt* (lit. *eyes [are] blue* vs. *blue eyes*, i.e. frightened), etc. It can be seen that in these cases, the change in order, from "body part + adjective" to "adjective + body part" in Vietnamese, is a means to express the modality stance, when these combinations are used in communication. According to the heteroglossia approach, they are means for the speaker/writer to express his or her attitude, i.e. a kind of modality meaning, in opposition to or agreement with the opinion of the interlocutor. Also, the change in order to express the stance will be the material for speakers to use in heteroglossic utterances, either dialogistic expansion or contraction**.**

3.2. Change of word order in sentences and the grammaticalization process in Vietnamese

The heteroglossia approach also contributes to clarifying some aspects of the process of grammaticalization in Vietnamese, forming final modal particles, and showing the speaker's attitudes in a discourse to the listener. Grammaticalization is a process of changing content words and function words in sentences, and it is a common tendency in many languages. In Vietnamese, content words can be grammaticalized to become function words, as we demonstrate in Nguyễn Văn Hiệp (2008) where several modal final particles are shown to be derived from content words in Vietnamese, which partly reflects the development of Vietnamese, a typical isolating language. Here we return to 7 of such modal final particles formed by the process of grammaticalization, but analyze them in terms of the change in word order, which make utterances heteroglossic.

Seven modal final particles *mất, thật, nghe, xem, đây, đấy, đi,* among many others in Vietnamese, are originally content words or demonstrative pronouns, but changes of word order result in changes of their meanings and functions in sentences. It is easy to see that utterances containing these final modal particles are only understandable in the context of dialogues where different points of view exist.

Mất (lose)

As a main verb, *mất* means 'lose, do not have, do not see, do not exist (temporarily or permanently)', e.g. ***Mất*** *tín hiệu liên lạc* (lost connection/ disconnected) (Hoàng Phê, 1996, p.601). Nevertheless, as an additional modal particle, ***mất*** is put after a verb to express an aspectual meaning that can be called "resultative aspect" as in the following examples,

(26) Cái xe này, tôi phải mua **mất** sáu trăm triệu.

'This vehicle, I had to buy at **as much as** 600 million.'

(27) Tôi đợi **mất** 2 tiếng mới có xe buýt.

'I spent **as long as** 2 hours waiting for the bus.'

In another position – at the end of a sentence, ***mất*** indicates a negative view (unexpected or worrying) of the speaker on a situation that may occur. For instance,

(28) Trời như thế này thì mai mưa **mất**!

'If the weather is like this, it'll rain tomorrow.'

(29) Học như thế này thì cô ta ở lại lớp **mất**!

'If she learns/continues to learn like this, she'll be retained.' (i.e. have to repeat the grade)

When ***mất*** is used to convey aspectual meanings as in examples (26) and (27), these meanings are still obviously related to its orginal conceptual meaning (its primary meaning as a content word in the opposing pair *được(get)/mất (lose*). However, when ***mất*** plays the role of a final modal particle as in (28) and (29), it expresses the speaker's view on the state of affairs in those sentences.

Thật (truly)

Thật stands after a noun as a content word with the following basic meaning: it truly refects an actual concept or a name, not a fake one, e.g. *Hàng* ***thật*** (**authentic** products) (Hoàng Phê, 1996, p.895). However, ***thật*** at the end of a sentence expresses an acceptance or an assertion of the speaker to the state of affairs in a sentence when the speaker wants to propose a new thought that is opposite to his old one. For example,

(30) Tiền này là tiền giả **thật**!

'[It dawns to me that] This is **truly** fake money.'

(31) Cô ta không thể tranh luận được **thật**!

'[I now acknowledge that] She can't make an argument.'

Xem (see)

As a content word, ***xem*** is a verb to express the basic meaning *see* (with one's eyes), e.g. ***Xem*** *phong cảnh* (see a beautiful view) (Hoàng Phê, 1996, p.1107). When ***xem*** is placed at the end of a sentence, it implies that the speaker wants the listener(s) to do (or get involved in) the action mentioned in the sentence. According to Searle (1969), ***xem*** is considered one of the speech act markers in directives in this meaning, as in

(32) Ăn quả nho này thử **xem**! Thấy vị có giống nho Mỹ không?

'Eat this grape, please. (**See if**) It tastes like American grapes?'

(33) Nghe thử bài hát này **xem**! Có vẻ được đấy!

'Listen to this song, please. (**See if**) It seems good.'

(34) Ngửi cái này **xem**! Hình như có mùi oải hương?

'Smell this. (**See if**) It smells like lavender?'

Đi (go)

As a content word, ***đi*** is a verb which means self-moving by successive movements of legs, with one foot always touching the ground while the other one is raised and brought to a distance, e.g. *Trẻ* ***đi*** *chưa vững* (The child doesn't walk well/ still toddles) (Hoàng Phê, 1996, p.301). When ***đi*** is placed at the end of a sentence, it means the speaker wants the listener to do the action mentioned in the sentence, and it is also considered a speech act marker in directives, e.g.

(35) Đánh, đánh bỏ mẹ cái thằng mèo nhép kia đi!

'Hit, hit the shit off that petite cat!'

(36) Kìa, mình ăn đi. Có chịu khó ăn mới chóng khoẻ chứ!

'Please, try to eat a little bit. You need to eat more to get well soon!'

Nghe (listen/hear)

Nghe stands before a noun, a verb or occurs in a subject-verb clause as a content word with the meaning listen or hear, e.g. ***Nghe*** *có tiếng gõ cửa* (I heard someone knocking the door) (Hoàng Phê 1996, p.653). When ***nghe*** is placed at the end of a sentence, it expresses the speaker's expectation to a state of affair in the sentence. ***Nghe*** is also considered a speech act marker in directives like ***xem*** and ***đi***, as in

(37) Đi coi phim nhớ về sớm **nghe**!

'Remember to come home right after the movie, **ok**?'

(38) Cẩn thận đừng vào bếp đang nấu **nghe**!

'Be careful not to touch the cooking stove, **please**!'

Đây/Đấy (here/there or this/that)

Đây and **đấy** are deixes that point to a place. ***Đây*** is used to indicate a place near the speaker, and ***đấy*** shows a further place. When these two words are used as metonyms, they can refer to people or animals present in a scope of space. For example,

(39) **Đây** là thầy Nam.

'This is teacher Nam.'

(40) Đây là cậu tôi, còn đấy là dì tôi.

'This is my uncle, and that is my aunt.'

When occurring at the end of a sentence, these deixes work as final modal particles with extensive meanings to convey a forecast of moments of actions. Compare:

(41) Tôi đi về nhà **đây**.

'I am leaving for home now.' [I hereby inform you that]

(42) Tôi đi **đấy**.

'I'll go [you know].'

(41) is considered a statement that the speaker is going to perform the action of going promptly while (42) can be understood as a verbal warning that the action *go* may happen.

Đây/đấy also express epistemic modality, which indicates the speaker's assertion about the truth of a proposition in accordance with present or past evidences in terms of time. ***Đây*** is used to express speaker's assertion upon present evidences that the speaker is experiencing at the speaking time. For instance, when we first see a young boy playing the guitar, we might say,

(43) Cậu này chơi ghi ta được **đây**.

'This guy may play the guitar well.'

If we have ever watched and/or listened to that guitarist before, we might say,

(44) Cậu này chơi ghi ta được **đấy**.

'This guy plays the guitar well.'

Therefore, as a final modal particle, ***đấy*** expresses an assertion upon past evidences before the speaking time.

As mentioned above, from SFG standpoint, word order change is a means to make meaning; in other words, word order differences provide choices for expressing meanings, including experiencial meaning, interpersonal meaning and textual meaning. In many cases, in a sentence, a change in the word order can lead to a simultaneous change of not just one, but all those meanings. For instance, the word đây exerts its different meanings and functions when it is put in different positions in a sentence, resulting in a very interesting sentence like the following:

(45) Đây$_1$ đi đây$_2$ đây$_3$!

'I have to go now!'

In 45), đây$_1$ acts as a first person pronoun and the subject of the sentence whereas đây$_2$ is a demonstrative pronoun in the role of a complement; and đây$_3$ is a final modal particle (see Nguyễn Minh Thuyết & Nguyễn Văn Hiệp, 1998; Nguyễn Văn Hiệp, 2009 for the concept of final modal particle), which indicates that an action is going to be carried out. It can be seen that word order change that leads to the grammaticalization of content words, i.e. content words turning into final modal particles, makes utterances heteroglossic.

4. Conclusion

To sum up, like many other languages, word order plays a very important role in the Vietnamese language. A general principle is changing word order means changing forms of expression, and leads to changes of meanings and functions. Among various approaches to word order, the heteroglossia approach shows that a change in word order is a change of expression to express different types of modal meanings in dialogues that contain a multitude of differing views. That indicates that the same material or meaning-making resourse can be used in different contexts with different functions.

References

Vietnamese

Lê Đông, Nguyễn Văn Hiệp (2003). Khái niệm tình thái trong ngôn ngữ học. *Ngôn ngữ*, Số 6 và Số 7-2003.

Đinh Văn Đức (1986). *Ngữ pháp tiếng Việt*. Hà Nội: Nxb Giáo dục.

Cao Xuân Hạo (1991). *Tiếng Việt - Sơ thảo ngữ pháp chức năng*, tập 1. Tp Hồ Chí Minh: Nxb Khoa học Xã hội.

Cao Xuân Hạo (1998). *Tiếng Việt - mấy vấn đề ngữ âm, ngữ pháp, ngữ nghĩa.* Hà Nội: Nxb Giáo dục.

Nguyễn Văn Hiệp (2001a). Hướng đến một cách miêu tả và phân loại các tiểu từ tình thái cuối câu tiếng Việt. *Ngôn ngữ*, số 5-2001.

Nguyễn Văn Hiệp (2001b). Về một khía cạnh phân tích tầm tác động tình thái. *Ngôn ngữ*, số 11/2001, tr. 42-50.

Nguyễn Văn Hiệp (2004). Về một khía cạnh phát triển của tiếng Việt (thể hiện qua hiện tượng ngữ pháp hóa hình thành một số tiểu từ tình thái cuối câu). *Ngôn ngữ*, số 11-2004.

Nguyễn Văn Hiệp (2008). *Cơ sở ngữ nghĩa phân tích cú pháp*. Hà Nội: Nxb Giáo dục.

Nguyễn Văn Hiệp (2009). *Cú pháp tiếng Việt*. Hà Nội: Nxb Giáo dục Việt Nam.

Nguyễn Văn Hiệp (2018). Các tiểu từ tình thái cuối câu trong tiếng Việt nhìn từ góc độ "dị thanh". Kỷ yếu hội thảo khoa học quốc tế "Ngôn ngữ học Việt Nam - những chặng đường phát triển và hội nhập quốc tế". Đại học Sư phạm Đà Nẵng và Viện Ngôn ngữ học: Nxb Thông tin và Truyền thông.

Nguyễn Văn Hiệp (2019). Nghiên cứu quán ngữ tình thái từ cách tiếp cận dị thanh. *Khoa học Xã hội Việt Nam*, số 7/2019.

Hoàng Phê (chủ biên) (1996). *Từ điển tiếng Việt* (Vietnamese Dictionary). Viện Ngôn ngữ học: NXB Đà Nẵng.

Nguyễn Anh Quế (1988). *Hư từ trong tiếng Việt hiện đại*. Hà Nội: Nxb Khoa học Xã hội.

Nguyễn Kim Thản (1964). *Nghiên cứu về ngữ pháp tiếng Việt*, Tập 2. Hà Nội: Nxb Khoa học Xã hội.

Lý Toàn Thắng (2004). *Lý thuyết trật tự từ trong cú pháp*. Hà Nội: Nxb Đại học Quốc gia Hà Nội.

Nguyễn Minh Thuyết & Nguyễn Văn Hiệp (1998). *Thành phần câu tiếng Việt*. Hà Nội: Nxb Đại học Quốc gia Hà Nội.

Hoàng Tuệ (1988). Về khái niệm tình thái. *Ngôn ngữ*, số phụ 1/1988.

Phạm Hùng Việt (2003). *Trợ từ trong tiếng Việt hiện đại.* Hà Nội: Nxb Khoa học Xã hội.

English

Bakhtin, M.M. (1935 [1982]). *The Dialogic Imagination: Four Essays*. Austin: University of Texas Press.

Halliday M.A.K. (1985). *An introduction to Functional Grammar*. London: Arnold.

Lyons, J. (1977). *Semantics*, vol 2. Cambridge: Cambridge University Press.

Lyons, J. (1995). Linguistic Semantics - An introduction. Cambridge: Cambridge University Press.

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

Palmer, F.R (2001). *Mood and Modality.* Cambridge: Cambridge University Press.

Searle, J.R. (1969). *Speech acts: an essay in the philosophy of language*. Cambridge: Cambridge University Press.

Traugott, E.C. and Heine, B. (eds) (1991). *Approaches to grammaticization*, Volume I. Amsterdam/Philadelphia: John Benjamins publishing company.

Voloshinov, V.N. (1995). *Marxism and the Philosophy of Language*. Translated by L.Matjka and I.R Tirunik. London: Routledge.

White, P.R.R. (2003). Beyond modality and hedging: A dialogic view of the language of intersubjective stance. *Text - The Hague Then Amsterdam Then Berlin*, 23(2), 259-284.

White, P.P.R. and Motoki, S. (2006). Dialogistic positioning and interpersonal style - A framework for stylistic comparison. In *Pragmatic markers in contrast*, eds by Karin Aijmer; A M Simon-Vandenbergen, Elsevier.

DỊ THANH: MỘT CÁCH TIẾP CẬN DỰA TRÊN NGỮ PHÁP CHỨC NĂNG HỆ THỐNG ĐỂ NGHIÊN CỨU TRẬT TỰ TỪ VỚI TƯ CÁCH PHƯƠNG THỨC BIỂU THỊ TÌNH THÁI TRONG TIẾNG VIỆT

Nguyễn Văn Hiệp

Viện Ngôn ngữ học Việt Nam
Số 9 Kim Mã Thượng, Ba Đình, Hà Nội, Việt Nam

Tóm tắt: Bài viết này trước tiên giới thiệu cách tiếp cận dị thanh, một sự phát triển của Ngữ pháp chức năng hệ thống (SFG) trong khung lí thuyết Ngôn ngữ học chức năng hệ thống theo Halliday. Sau đó, chúng tôi sẽ áp dụng cách tiếp cận dị thanh để nghiên cứu vai trò của trật tự từ trong tiếng Việt, một ngôn ngữ đơn lập điển hình. Sự thay đổi trật tự từ được xem là cách biểu đạt tình thái, như có thể thấy qua một số ví dụ thú vị trong tiếng Việt, điều đó chứng tỏ ngữ pháp chức năng hệ thống và cách tiếp cận dị thanh với tư cách một dạng phát triển của nó thật sự là cách tiếp cận hiệu quả để nghiên cứu trật tự từ trong tiếng Việt.

Từ khóa: ngôn ngữ đơn lập, trật tự từ, tình thái, Ngữ pháp chức năng hệ thống, cách tiếp cận dị thanh, ngữ pháp hóa

TESTS AS DRIVERS OF CHANGE IN EDUCATION: CONTEXTUALISING WASHBACK, AND THE POSSIBILITY OF WASH-FORWARD

Gordon Allan*

British Council, 1-2 Kagaruzaka, Shinjuku-ku, Tokyo, 162-0825, Japan

Received 26 February 2020
Revised 15 March 2020; Accepted 25 July 2020

Abstract: Whilst the primary aim of tests is to measure ability, it is not uncommon for tests to be deployed in education systems with the intention, at least in part, of driving change in educational practice by making demands that teachers and learners are expected to meet. Washback is one way by which teaching and learning practices may adapt to a new test, but it is not the only possibility and often fails to occur as intended. This paper seeks to draw together ideas from different sources to place washback in the context of other possibilities. The concepts of adaptive implementation and programmed implementation are taken from Henrichsen's hybrid model of the diffusion/implementation of innovation in education systems. Washback is shown to act in parallel to but distinct from programmed implementation. The picture is completed with van Lier's concept of wash-forward, first outlined in 1989 but subsequently neglected in the literature. Wash-forward is illustrated with an example from the implementation of the National Matriculation English Test (NMET) in China. The intention is to provide an easily visualised, refreshed and more complete perspective on the processes operating when a new test is introduced as part of a strategy aimed at driving changes in teaching and learning practices; a scenario which is very relevant to the current movement towards four-skills English testing in East Asia and around the world.

Keywords: Washback, wash-forward, programmed implementation, innovation in education, exam reform

1. Introduction: Tests as drivers of change in education

High-stakes tests are often introduced in the hope of driving educational change by eliciting positive washback (for examples see Wall & Alderson, 1993; Cheng, 2002; Qi, 2005; MEXT, 2014). In such cases, it is hoped that the demands of a new test will help to drive changes in teaching and learning practices to produce more effective learning outcomes. The literature, however, shows that the desired washback often fails to occur (e.g. Wall & Alderson, 1993; Qi, 2005). It is therefore worth considering the mechanism by which tests might drive educational change, and what might be happening when the desired washback does not transpire.

The present paper first briefly reviews the concept of washback, then attempts to contextualise it by relating it to Henrichsen's hybrid model of the diffusion/implementation of innovation in education systems (Henrichsen, 1989). Finally, the concept of "wash-forward" (van-Lier, 1989, p.494) is added to make up a picture of two pairs of parallel processes acting from opposite sides to bridge the gap between the demands of a new test on one side and existing teaching and learning practices on the other.

* Tel.: +81-90-9954-3184
Email: Gordon.Allan@britishcouncil.or.jp

2. Washback

At its simplest, washback (synonymous with backwash) has been defined as "the effect that tests have on learning and teaching" (Hughes, 2003, p.53). Buck (1988) views washback as "a natural tendency for both teachers and students to tailor their classroom activities to the demands of the test" (p.17), particularly high-stakes tests, noting that it "can be either beneficial or harmful". Messick (1996) expands on this somewhat, defining washback as "the extent to which the introduction and use of a test influences language teachers and learners to do things they would not otherwise do that promote or inhibit language learning" (p.241). This "otherwise" is of course hypothetical, and usually gauged by a baseline study of what teachers and learners did prior to a test's introduction. Since this baseline may vary between contexts and/or individuals, the same test may elicit different washback from different contexts and individuals (examples of washback varying between individuals are summarised in table 1). Washback is therefore not a property of a test *per se*, but a consequence of test use.

Table 1: Examples of washback varying between individuals in the same context

Study	Alderson & Hamp-Lyons (1996)	Watanabe (1996)
Comparison	Language proficiency classes vs. TOEFL preparation classes	Preparation classes for the English sections of two types of examinations: Entrance examination for a national university [grammar-translation (GT) oriented] Entrance examination for a private university [non-GT oriented]
Context	Specialised language institute for foreign students in the USA "whose students are regularly admitted to USA universities" (p. 283)	*Yobiko* (privately-run extra-curricular entrance exam preparation school) for Japanese students in central Tokyo
Differences in washback reported	Study observed two teachers in TOEFL prep and regular lessons. Both teachers showed some consistent differences between their TOEFL-prep and regular classes, e.g. greater teacher talk time, fewer opportunities for student talk, and less use of pairwork. However, the two teachers' TOEFL-prep lessons also varied considerably in approach, e.g. use of metalanguage, test-practice in class time, and opportunities for discussion. In contrast to the two observed teachers, another teacher (new to TOEFL teaching, not observed) reported thinking that interaction was vital in a TOEFL prep class and described a number of interactive techniques she used in her TOEFL prep lessons.	Two teachers observed teaching exam preparation classes for both types of exam. One focused on translation and explaining structures for both exams, regardless of whether the exam was GT-oriented. The other varied his approach, translating and explaining more in the GT-oriented exam preparation lessons than in the non-GT-oriented exam preparation lessons.

More recently, Cheng (2005) defines washback as “an intended or unintended (accidental) direction and function of curriculum change on aspects of teaching and learning by means of a change of public examinations”, which is consistent with previous definitions in that washback refers to examinations changing teaching and learning practices, and not the reverse. Since washback has the potential to bring about change in educational practice, it is worth considering how it fits with a model of innovation in education.

3. Henrichsen’s hybrid model of diffusion of innovation in education

Henrichsen’s (1989) hybrid model of the diffusion/implementation of innovation in education systems follows the course of an innovation, from its antecedent state, through the process of its implementation, to the decision to adopt or reject and subsequent consequences. It remains a rich and comprehensive account of the factors influencing the success or otherwise of educational innovations.

Although washback has a role in the diffusion of innovation, it is not immediately obvious where it fits into Henrichsen’s (1989) model. It is a consequence of test use but does not appear in the consequences phase of the model because the model is concerned with the fate of the test (innovation) itself rather than the test’s effect on other practices. Instead, interaction between the test and its user system is considered as part of the process phase, in the section on “inter-elemental factors”.

Henrichsen (1989) argues convincingly that mismatch between an innovation and its user system potentially threatens the successful adoption of the innovation. Where tests are deployed to drive change in educational practices, however, the motive force for change can only be generated by deliberately creating precisely such a mismatch, since a test that is well-aligned with existing practice provides no motivation for change. When a test makes demands that are not being met by the educational status quo, a gap opens between the test on one side and teaching and learning practices on the other. To bring the two back into harmony, this gap must be closed.

A full account of the nature of potential dissonance between the demands of a new test and existing educational practice is beyond the scope of the present paper, which seeks to deal with the topic at a conceptual level. As illustrated in the examples summarised in table 2, however, in general terms it is not uncommon for new tests to emphasise the practical use of English for communicative purposes where this is perceived to be deficient in existing English teaching and learning practices.

Table 2: Examples of dissonance between the demands of a new test and existing educational practices, with the intention of inducing change in education systems

Study	Wall & Alderson (1993)	Qi (2005)
Context	Introduction of a new O-level English exam, custom-made to reinforce new textbooks, in Sri Lanka, 1988.	Introduction of the National Matriculation English Test (NMET) in China, 1985.

Nature of dissonance between the demands of the test and existing teaching and learning practice	Prior to this, teaching tended to be structurally based and focused on general reading skills, with little opportunity for communication. The new O-level and accompanying textbooks introduced greater emphasis on reading and writing for a communicative purpose, a previously neglected aspect of both skills.	Prior to the NMET, English teaching was characterised by a focus on grammatical structures and rote-learning of vocabulary, with little attention to communicative skills. The NMET shifted the focus from formal linguistic knowledge to the practical use of language.
Intended change	The development of more practical English communication skills.	"…to replace the time-honored traditional method in China's ELT with the widely accepted communicative approach." (p.145)

Henrichsen (1989, p.92) cites Roberts-Gray & Gray (1983) to describe two processes for the resolution of dissonance between an innovation and its user system: adaptive and programmed implementation. Adaptive implementation involves altering the innovation to fit the users, and may include such processes as localising test content, for example to ensure that pictures used in the test are recognisable to test-takers. Programmed implementation works in the opposite direction, seeking to change the user system to accommodate the innovation, an example of which might be a teacher training programme aimed at helping teachers to prepare students for a new test. Table 3 shows illustrative examples of programmed and adaptive implementation in practice.

Table 3: Examples of programmed and adaptive implementation from the Sri Lanka study (Wall & Alderson, 1993)

Study	Wall & Alderson (1993)
Context	Introduction of a new O-level English exam in Sri Lanka, 1988.
Adaptive implementation (aimed at changing the test to accommodate the test user system)	• The new exam was custom made to align with new textbooks. Test developers analysed the textbooks and drew up test specifications in consultation with the textbook writers. The new textbooks and new exam were therefore well-aligned and mutually supportive of each other. • There was a conscious attempt to choose texts, topics, and authentic tasks that were relevant to Sri Lankan school children and their likely purposes for using English. • Plans to employ continuous assessment to test oral skills were dropped because they proved practically and politically impossible.
Programmed implementation (aimed at changing the test user system to accommodate the new test)	• The first textbook in the new series was accompanied by a Teacher's Guide including guidance on the essentials of the approach teachers were expected to follow and how to use the new textbook material in the classroom. • Teacher training efforts also accompanied the introduction of the new textbooks.

Described thus, both adaptive and programmed implementation are active interventions to support the success of an innovation. Washback is distinct in that it is not an active intervention, but a consequence of a test's interaction with its educational context. Although washback acts in the same direction as programmed implementation (i.e. the users adapt to the innovation), it arises from teachers and learners as a response to the test. This view is consistent with that of Messick (1996), who is careful to maintain a distinction between washback and other drivers of change, such as training programmes or new textbooks. In addition, washback may be unpredictable and inconsistent (e.g. Watanabe, 1996) in contrast to what is implied by the term "programmed".

The relationship between programmed implementation and washback is an important one. As Cheng (2002) notes, while tests may provide the motivation for change, they do not provide the knowledge or skills required to enact that change. As teachers and learners respond to the demands of the test, any new knowledge or skills that may be required of them must be drawn from somewhere. A test can only elicit what teachers and learners have the capacity to provide. This may help to explain findings such as Wall and Alderson's (1993) observation that washback was evident in content taught but not teaching methodology.

The role of programmed implementation in creating the potential for positive washback is therefore crucial, for example via training to foster the development of new skills, or the development and provision of appropriate resources such as textbooks. As Cheng (2002) puts it:

The change to a new exam has informed teachers about what they might do differently, but it has not shown them how to do it. The washback effect can be fully realized only when all levels of organizations in the educational system are involved. In this sense, there must be a genuine involvement of educators and textbook writers. A change in the final examination alone will not achieve the intended goal (Cheng, 2002, p.108).

In summary, it seems helpful to consider washback as parallel to but distinct from programmed implementation. This raises the question of whether there is any equivalent parallel to adaptive implementation. The literature suggests that there is.

4. "Wash-forward" (van Lier, 1989)

4.1 What is wash-forward?

In his 1989 paper on the oral proficiency interview (OPI), Leo van Lier observes that:

By pushing for innovative testing techniques, particularly the OPI, proficiency advocates hope that a desirable washback effect will be created; in other words, if teachers and learners know that tests will demand both communicative ability and accuracy, the methodological focus of classroom work will change accordingly. This is something of an act of faith, of course, since it is also possible that classroom practices will prove so recalcitrant that they will force the OPI to shift in the direction of standard curricula. (van Lier, 1989, p.491)

The paper goes on to introduce the concept of "a possible 'wash-forward' (as opposed to washback) effect of methodological and curricular concerns carrying over into the rating" (van Lier, 1989, p.494). In other words, existing perceptions and/or practices may influence how a test construct is operationalised.

The example of possible wash-forward given by van Lier (1989) is that, in practice, aside from their training and the descriptors given in rating scales, examiners may tend to over-rely on linguistic features that are given prominence in syllabuses (e.g. conditional forms), particularly when making decisions at the boundaries between bands/grades. Relating this to Henrichsen's (1989) hybrid model, wash-forward acts in the same direction as adaptive implementation by "modifying or adapting the innovation so that it is more easily assimilated into user practices and values" (Roberts-Gray & Gray, 1983, p.216).

It seems obvious that van Lier's (1989) concept of wash-forward must be a possibility, but apart from a brief mention by Johnstone (1994), it does not appear to have been followed up in the literature. It is not immediately obvious why not. When this paper was first presented at the AALA conference in Hanoi, 2019, it was suggested that perhaps wash-forward had been considered as a type of washback. This may be the case, since both washback and wash-forward are consequences of test use, but it goes against the definition of washback as "the effect that tests have on learning and teaching" (Hughes, 2003, p.53), consistent with other definitions (e.g. Buck, 1988; Messick, 1996; Cheng, 2005). Thus defined, washback cannot include its opposite, i.e. the effects that teaching and learning have on tests, so a distinct term is required, hence wash-forward.

"Wash-forward" (van Lier, 1989) is perhaps not a particularly intuitive term for the phenomenon it refers to. It was presumably chosen because it sounds like the opposite of washback, but a more clearly descriptive term may well be preferable, and it is even possible that the lack of such has played a role in the subsequent neglect of the concept. To avoid confusion, however, no alternative terms are suggested here, and wash-forward is used throughout.

It is also worth noting that the suggested mechanism for wash-forward in van Lier's (1989) example is via the rating process. This mechanism can only influence assessment of the productive skills (speaking and listening), since there is no equivalent rating process for the receptive skills (listening and reading). Further research would be required to determine whether there is any other mechanism by which wash-forward might occur.

Despite the term and concept being neglected, however, if wash-forward is a real possibility we might expect examples of it to be found in the literature. Qi's (2005) account of the National Matriculation English Test (NMET) in China appears to document just such a possible occurrence in practice.

4.2 A possible example of wash-forward in the implementation of the NMET, China (Qi, 2005)

Introduced in 1985, one of the aims of the English component of the NMET was to promote a shift away from traditional methods of English education towards a more communicative approach (Qi, 2005). As part of this effort, a communicative context (audience, purpose) was provided in the rubric for the writing task, and the marking criteria included appropriacy (Qi, 2005). However, the inclusion of appropriacy only lasted for the first six years, until 1990, before it disappeared from the marking criteria, subsequently reappearing only inconsistently and/or indirectly.

In her study, Qi found that teachers and students did not see the specified communicative context as a basis upon which to choose an appropriate writing style. Instead, they "interpreted the trait measured … as the ability to extract the prescribed content from the input and put it down neatly in grammatically correct English sentences" (Qi, 2005, p. 157). In addition, NMET markers' advice about improving scores, tended to emphasise handwriting, accuracy and the extraction of key points from input, but did not mention appropriacy (Qi, 2005).

Qi (2005) concluded that "appropriacy does not significantly affect the scores for writing in the NMET" (p.158), an apparent example of construct under-representation (Messick, 1996), not by design but due to the way the construct was operationalised. Qi (2005) attributed this in part to teachers lacking sufficient experience of English communication to judge appropriacy. She cites an example of a teacher recommending over-polite language as appropriate for a letter to a railway station lost property office: "Anyway, you should sound polite… It's better if you write 'I would be very much obliged if you could kindly help me to find it.'" (Qi, 2005, p. 159). This is consistent with concerns raised by Cheng (2002) that for change to be enacted as desired, those enacting it require the knowledge and skills to do so.

So, despite the inclusion of appropriacy in the early marking criteria, it seems questionable whether that aspect of the construct was ever operationalised. In the absence of a sufficient base of the required knowledge and skills, it appears that educational practices may have proved "so recalcitrant" (van Lier, 1989, p.491) as to cause an effective alteration of the construct of the test; wash-forward as opposed to the intended washback.

Qi's (2005) account reminds us that a high-stakes test is not an immovable object but is subject to the influence of the context(s) within which it is deployed.

5. Summary

The above has sought to place washback in the broader context of the interactions that may occur when a test is deployed with the aim of driving change in teaching and learning. A schematic representation of the four processes described is given in figure 1:

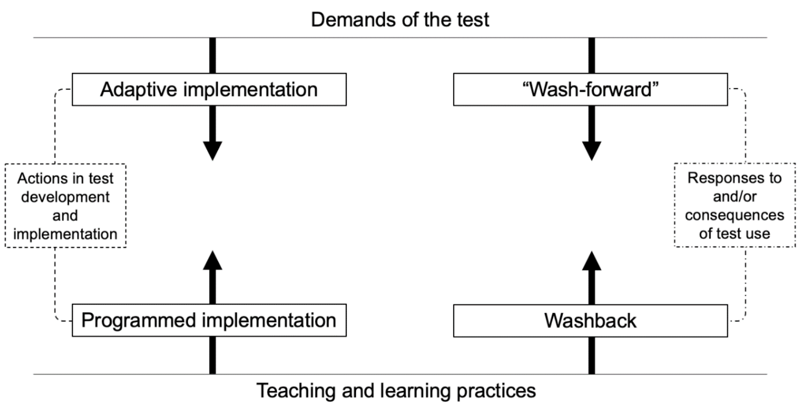


Figure 1: The processes by which dissonance between the demands of a test and the practices of teachers and learners might be resolved

The processes of programmed and adaptive implementation cited by Henrichsen (1989) are complemented by the well-established concept of washback on the one hand, and van Lier's (1989) concept of wash-forward on the other. Programmed and adaptive implementation are considered as actions (interventions) taken by those implementing a new test, while washback and wash-forward are responses to and/or consequences of test use. It should be noted that, as with other consequences of innovation, washback and wash-forward may be immediate and/or delayed, direct and/or indirect, manifest and/or latent, functional and/or dysfunctional (Henrichsen, 1989, p.95).

6. Implications

O'Sullivan (in press) argues convincingly that for an education system to work, assessment must be in harmony with the curriculum and all aspects of delivery. This is consistent with Henrichsen's argument that mismatch between an innovation and its user system is a threat to the success of the innovation. O'Sullivan's argument takes this further; in a comprehensive learning system, any mismatch between the key elements of assessment, curriculum, and delivery does not only threaten the mismatched element, but puts the whole system at risk (O'Sullivan, in press).

The present paper has sought to examine in detail one possible source of mismatch within a learning system, and place the widely studied phenomenon of washback in the context of other processes that may operate when such a mismatch occurs. Since the different parts of an effective learning system are interconnected, there are implications for a number of different stakeholder groups. Central to these is the importance of communication between various stakeholder groups, in terms that everyone can understand.

6.1 Implications for test developers

Test developers need to work with curriculum planners and classroom practitioners to achieve a harmonious system. They need to help inform teachers and learners about tests, but also need to research and understand the context in which the test will be used. Such understanding can inform adaptive implementation, bringing the test closer to the needs and potential of teachers and learners. It can also be helpful in identifying opportunities for programmed implementation to equip teachers with the knowledge and skills to help learners meet the demands of any new test, thereby helping to generate potential for positive washback.

6.2 Implications for education policy makers

Qi's (2005) account of the NMET makes clear that if a test is not well-aligned with the knowledge and abilities of those implementing it, wash-forward is a possible consequence, which may undermine the test construct and its potential to drive desired learning outcomes. Given the expense involved in developing and implementing a high-stakes test, this might represent a very costly failure. Policy makers therefore need to work with test developers, teachers, and researchers to understand the demands of any new test, the capacity of teachers and learners to meet these demands, and the support required to make the system work.

6.3 Implications for teachers and learners

Teachers and learners also share some responsibility for understanding assessment and how it relates to classroom practice. It is important that they understand what is being tested so that they can approach language learning and test preparation constructively. Developing the knowledge and skills required may involve engaging with training, asking questions, and being open to new ways of doing things. Equally, it may involve working with researchers to help them understand how a test is perceived and identify what support may be needed to make it work in context. Just as it is important for test developers and policy makers to listen, it is equally important for teachers and learners to have a voice in the conversation, and to use it to feed back their perceptions and experience.

6.4 Implications for researchers

Researchers have an important role to play in facilitating effective communication between the elements of a learning system, by collecting and analysing data including different stakeholder perspectives. It is therefore important that researchers understand the interactions between the elements of the system.

The contextualisation in the present paper is a reminder that washback is not synonymous with the broader category of impact, but one aspect of impact. The relatively high profile of washback in language testing research has perhaps put it at risk of becoming a catch-all term occasionally (mis)used to refer to consequences of test use that are not covered by the clear and consistent definitions in the literature. For example, a recent review of "ten empirical studies of washback in language teaching conducted around the world between 2011 and 2018" (Ahmmed & Rahman, 2019, p.153), whilst an interesting and useful contribution to the literature, included three studies that do not appear to be concerned with washback as defined. Of these three, two were experimental in design (Khoshsima, Saed & Mousaei, 2018; Khodabakhshzadeh, Zardkanloo & Alipoor, 2017) and one involved action research aimed at determining what might constitute effective test-preparation (Munoz, 2017). Unlike experimental and action research studies, however, washback is not an intervention, but is what teachers and learners are observed to do in response to a test. This is not always what we might recommend them to do, and it is helpful to preserve a distinction between the actual, the desirable, and the experimental.

Studies of effective test-preparation strategies are of interest in their own right, and may help to inform both test development and programmed implementation, but they are not studies of washback per se. An understanding of the processes which operate alongside washback may help researchers to more clearly contextualise research and avoid blurring the established concept of washback.

7. Conclusion

The present paper has sought to bring together strands from different sources to present a fresh way of looking at the role that tests might play in driving educational change by contextualising the processes involved and how they relate to each other. Although the ideas presented are not new, at least one of them has been neglected for decades and it is suggested that without wash-forward the bigger picture is incomplete. The demands of a new test may provide the motivation for change, but interventions such as training programmes, and appropriate textbooks (i.e. programmed implementation) have an important role to play in increasing the potential for high-stakes tests to generate positive impact. Failing to equip teachers and learners with the knowledge and skills required of them may mean that they have little option but to continue doing what they know, and this may even lead to the test being forced to change (wash-forward).

The visualisation presented in figure 1 may be of use to students and researchers seeking to understand washback and related ideas, but it is perhaps of greater potential value in communicating with non-specialists in assessment, such as policy makers who have the power to make decisions about new tests and their implementation. Without over-simplifying ideas, it is essential that the assessment community is able to communicate other educational stakeholder groups in ways that non-specialists can understand and integrate with their existing knowledge. It is hoped the present paper is a helpful contribution to that effort.

Acknowledgements

My thanks to Dianne Wall for her insight and help while supervising the dissertation that sparked the thinking outlined in this paper, and to my colleagues at the British Council, especially Barry O'Sullivan, Robin Skipsey, Chie Yasuda, Mina Patel, Sheryl Cooke, Richard Spiby, and Jamie Dunlea, for their comments and questions in discussion as the idea for the paper took shape. I would also like to thank the anonymous reviewer whose insightful comments were very helpful in improving the final draft. Any remaining deficiencies are entirely my own.

References

Ahmmed & Rahman, (2019). Review of empirical studies of washback in language testing. *Bangladesh Maritime Journal, 3*(1), 150-162.

Alderson, J.C., & Hamp-Lyons, L. (1996). TOEFL preparation courses: a study of washback. *Language Testing, 13*, 280-297.

Buck G (1988). Testing listening comprehension in Japanese university entrance examinations. *JALT Journal, 10*, 15-42.

Cheng, L. (2002). The washback effect on classroom teaching of changes in public examinations. In Savignon, S.J. (Ed) *Interpreting communicative language teaching: concepts and concerns in teacher education.* Yale University Press

Cheng, L. (2005). *Changing Language Teaching Through Language Testing: A Washback Study*. Cambridge, U.K: Cambridge University Press.

Henrichsen, L. (1989). *Diffusion of Innovations in English Language Teaching: The ELEC Effort in Japan, 1956-1968.* Greenwood Press Inc., Connecticut, USA.

Hughes, A. (2003). *Testing for language teachers* (2nd ed.). Cambridge, U.K.: Cambridge University Press.

Johnstone, R. (1994). Research on language learning and teaching: 1993. *Language Teaching, 27*(3), 145-156. doi:10.1017/S0261444800007795

Khodabakhshzadeh, H., Zardkanloo, R., and Alipoor, I. (2017). The effect of mock tests on Iranian EFL learners' test scores. *International Journal of Education & Literacy Studies, 5*(3), 47-51.

Khoshsima, H., Saed, A., and Mousaei, F. (2018). Exploring the effect of teaching test-taking strategies on intermediate level learners on reading section of IELTS; learners' attitude in focus. *Advances in Language and Literary Studies, 9*(2), 4-9.

Messick, S. (1996). Validity and washback in language testing. *Language Testing, 13,* 241-56.

MEXT (2014). *Report on the future improvement and enhancement of English education (outline): Five recommendations on the English education reform plan responding to the rapid globalization.* Retrieved from http://www.mext.go.jp/en/news/topics/detail/1372625.htm

Munoz, R. E. (2017). The effect of washback on EFL learners' attitudes toward tests. *Studies in English Language Teaching, 5*(3), 516-530.

O'Sullivan, B. (in press). The comprehensive learning system. *British Council perspectives on English language policy and education.*

Qi, L. (2005). Stakeholders' conflicting aims undermine the washback function of a high-stakes test. *Language Testing, 22* (2), 142-173.

Roberts-Gray, C., & Gray, T. (1983). Implementing Innovations. *Knowledge: Creation, Diffusion, Utilisation, 5*(2), 213-232.

van Lier, L. (1989). Reeling, writhing, drawling, stretching, and fainting in coils: oral proficiency interviews as conversation. *TESOL Quarterly, 23*(3), *(Sep. 1989),* 489-508.

Wall, D., & Alderson, J.C. (1993) Examining washback: the Sri Lankan impact study. *Language Testing, 10,* 41-69.

Watanabe, Y. (1996). Does grammar translation come from the entrance examination? Preliminary findings from classroom-based research. *Language Testing, 13*, 318-33.

KIỂM TRA – ĐỘNG LỰC THAY ĐỔI TRONG GIÁO DỤC: XÁC ĐỊNH TÌNH HÌNH TÁC ĐỘNG DỘI NGƯỢC VÀ KHẢ NĂNG TÁC ĐỘNG THÚC ĐẨY

Gordon Allan

Hội đồng Anh, 1-2 Kagaruzaka, Shinjuku-ku, Tokyo, 162-0825, Nhật Bản

Tóm tắt: Mặc dù mục đích chính của các bài kiểm tra là đo lường năng lực thí sinh, nhưng không có gì lạ khi các bài kiểm tra được triển khai trong các hệ thống giáo dục với mục đích, ít nhất là một phần, để thúc đẩy sự thay đổi trong thực tiễn giáo dục bằng cách đưa ra yêu cầu mà giáo viên và người học được hi vọng có thể đáp ứng. Tuy nhiên tác động của kiểm tra đánh giá lên hoạt động dạy và học (tác động dội ngược - washback) không phải là điều duy nhất có thể xảy ra và cũng thường không xảy ra đúng như mong muốn. Nghiên cứu này thu thập các ý tưởng khác nhau nhằm đưa ra các khả năng khác. Mô hình đưa ra liên quan tới các khái niệm "điều chỉnh bài thi" (adaptive implementation) và "điều chỉnh hệ thống" (programmed implementation) được lấy từ mô hình của Henrichsen về các điều chỉnh mang tính đổi mới trong hệ thống giáo dục. Tác động của kiểm tra đánh giá lên hoạt động dạy và học (washback) được cho thấy diễn ra song song nhưng cũng riêng biệt với điều chỉnh hệ thống (programmed implementation). Bức tranh toàn cảnh được hoàn thiện với khái niệm của Lier về "wash-forward" – tác động của hoạt động dạy và học lên kiểm tra đánh giá, một khái niệm được đưa ra lần đầu tiên vào năm 1989 nhưng sau đó đã dần bị lãng quên. Tác động thúc đẩy hoạt động dạy và học lên kiểm tra đánh giá (wash-forward) được minh họa bằng một ví dụ từ việc triển khai Bài thi tiếng Anh quốc gia National Matriculation English Test (NMET) tại Trung Quốc. Ví dụ này nhằm đưa ra một góc nhìn dễ hình dung, mới mẻ, và đầy đủ hơn về các quá trình giới thiệu một bài thi mới trong chiến lược thúc đẩy những thay đổi trong thực tiễn dạy và học; một ví dụ rất phù hợp với phong trào hướng tới kiểm tra đánh giá bốn kỹ năng tiếng Anh hiện nay ở Đông Á và trên toàn thế giới.

Từ khóa: Tác động của kiểm tra đánh giá lên dạy và học, tác động của dạy và học lên kiểm tra đánh giá, điều chỉnh hệ thống, đổi mới giáo dục, cải cách thi cử

ASSESSING REAL-WORLD USE OF ENGLISH AS A LINGUA FRANCA (ELF): A VALIDITY ARGUMENT

Sheryl Cooke*

British Council, University of Jyväskylä, Finland
989 Beijing West Road, Shanghai, 200041, PR China

Received 26 February 2020
Revised 20 March 2020; Accepted 20 July 2020

Abstract: Real-world use of English involves speakers and listeners from various linguistic backgrounds whose primary goal is mutual comprehensibility and the majority of conversations in English do not involve speakers from the Inner Circle (Graddol, 2006; Kirkpatrick, 2007). Yet, rather than focusing on comprehensibility, many tests continue to measure spoken performance with reference to an idealised, native-speaker form, weakening the validity of these tests in evaluating authentic spoken communicative competence as it is used in a global lingua franca context and leading to a narrowing of the construct of ELF, or to the inclusion of construct irrelevant factors.

Validation of a test of English as a tool for global communication includes demonstrating the link between the construct (real-world communicative ability in a particular context) and the test tasks and rating criteria (McNamara, 2006), and evidence to support the interpretation of a test score needs to be presented as part of the overall validity argument. First, this paper argues that the context of English use that many high-stakes test-takers aspire to – that of English for Academic Purposes (EAP) – is frequently an ELF context; second, Toulmin's (2003) argument schema is leveraged to explore what evidence is required to support warrants and claims that a test provides a valid representation of a test-taker's ability to use ELF. The framework as it relates to the validation of language tests in general is presented and the model is then applied to two tests of spoken English by way of illustration. Although examples are included, the main aim is to provide a theoretical justification for a focus on comprehensibility and the inclusion of linguistic variation in the assessment of ELF and to present a validation framework that can be applied by test developers and test users.

Keywords: English as a lingua franca, test validity, comprehensibility

1. Introduction

"Speak English!" said the Eaglet. "I don't know the meaning of half those long words, and I don't believe you do either!"

— Lewis Carroll, 1865 in *The Adventures of Alice in Wonderland*

English has long been recognised as the major international language for communication across a range of different domains, from academic conferences to business negotiations, from aviation to the international space station, from the United Nations to popular culture. 'English', however, is a broad term that encapsulates a growing range of English*es*, from the forms spoken in the traditional English-speaking countries of the UK, North America and Australasia, to the now established varieties spoken in ex-British colonies such as Singaporean English, South

* Tel.: +8613818274299
Email: Sheryl.cooke@britishcouncil.org.cn

African English and Indian English, and extending further to the learning and active use of English by speakers from a myriad of different language backgrounds for the purpose of international communication.

Native-speaker (NS) – and particularly 'standard' forms of NS English such as Southern British English or general American English – have long held an elevated position in the teaching and learning of English. Quirk (1985) argued that a standard form served the needs of non-native English speakers (NNS) because their communicative purposes were 'narrow'. Others have attempted to defend or define Standard English: Williams (1980) with a focus on a standard form of US English and Peters (1995) with Australian English. Davies (1999) concluded that *the* Standard form could be considered the 'language of the educated'.

Berns (2006, p. 723-724) outlines the key assumptions behind NS norms:

"(1) everyone learning English does so in order to interact with native speakers;

(2) the communicative competence learners need to develop is the native speaker's; and

(3) learning English means dealing with the sociocultural realities of English or the US, that is, British or American ways of doing, thinking and being."

But evidence from the real-world use of English debunks these assertions: real-world communication in English involves speakers and listeners from various linguistic backgrounds whose primary goal is successful communication. Furthermore, almost 20 years ago, Crystal noted that only a quarter of the world's English language speakers are NS users (Crystal, 2003); Ethnologue (Paul, Simons & Fennig, 2020) puts the total number of users of English in all countries at 1,268,100,190 of which 369,704,070 use it as an L1 and 898,396,120 as an L2, with other words, 71% of the world's speakers of English are not L1 users of the language. Given the international function of English in the geopolitical, economic and academic spheres (Ammon, 2010), it is clear that the majority of conversations in English do not involve speakers from the countries in Kachru's (1985) well-known Inner Circle (Graddol, 2006; Kirkpatrick, 2007; Nelson, 2011). Learners of English do not, in the majority of cases, have as their goal conversation with native speakers, they do not need to develop native speaker proficiency to achieve their communicative goals, and the sociocultural reality they operate in is diverse, dynamic and more likely to involve cultural and social characteristics of the Chinese, Indians or Brazilians than someone from the Inner Circle. The communicatively successful use of ELF by millions of speakers from varied backgrounds occurs in a range of different domains, both personal and professional (Seidlhofer, 2011).

There has been increasing recognition of the need for language learning and teaching to reflect these realities and support for this has been voiced in the academic community – Jenkins (2000) and Seidlhofer (2011), amongst others. Galloway (2018) explicitly points to the importance of teaching learners to communicate in a global context. There is also a move towards more inclusive course books that reflect the sociocultural reality referred to above, e.g. MacMillan Global course books, and the inclusion of a variety of NS and NNS accents in listening texts, although to what degree these attempts go beyond surface-level recognition of the reality of English use is debatable (Galloway, 2018).

In the field of language assessment specifically, the question of which English to test is particularly pertinent given the consequences tests have on the lives of individuals, on society more generally, and the washback effect that a high-stakes test has on teaching and learning. If a test-taker is to be tested on a certain variety of English, or a narrow range of varieties, then that test-taker will focus on studying those varieties and seek out exposure to those forms, even where this does not reflect their current or future communicative context. In a chicken-and-egg situation, the varieties that students learn to prepare for the standard-form tests are then used to support continued testing of only standard forms of English because exposure to other forms is limited in the classroom and in textbooks.

The question of whether, and how, to reflect real-world use of English in tests has been discussed by, amongst others, Elder & Harding (2008), Jenkins (2006), McNamara (2014) and Harding & McNamara (2017). Graddol, too, explicitly mentions testing in relation to the new sociocultural reality of English language use: "The way English is taught and assessed should reflect the needs and aspirations of the ever-growing number of non-native speakers who use English to communicate with other non-natives" (Graddol, 2006, p. 87).

This paper presents a theoretical perspective to the challenge of assessing English as a lingua franca (ELF) and test validity is at the core of the discussion. Frameworks used for the evaluation of test validity are presented and Toulmin's (2003) argument framework is applied to the context of assessing ELF. I will present an argument that centres around two key assumptions:

- the first, central to the contemporary idea of validity in language assessment, is that *a test must reflect the real-world use of English in order to be valid.*
- the second, at the core of the study of the global use of English and the study of World Englishes, is that *the real-world use of English is not limited to standard forms of the language but includes variety.*

What follows from this is that in order to be a valid assessment, a test must be linked to the domain of use and, in order to demonstrate validity, must present evidence to support the claim that the domain is not only *represented* but *adequately represented* in the test. The sections of this paper that follow consider validity and the link to the intended domain of use, investigate the domain in which candidates taking high-stakes tests for the purpose of academic study in an English-speaking context are likely to function, and use Toulmin's argument structure as an example of how evidence may be sought, presented and evaluated. Finally, an illustration of how the framework might be applied to two language tests is presented and next steps are suggested.

2. Validity

There are various approaches to establishing the validity of a test or assessment system, some more theoretical than others (Messick, 1989; Kane, 2012). Common to many approaches is the investigation of what tasks the test-taker needs to engage in in the real-world situation in which they communicate or intend to communicate. That is, a test must have a demonstrable link to the context – or domain – in which the ability is or will be put to use.

Various scholars have highlighted the importance of the link between the test and the context of use. Mislevy & Yin's (2012) work on *Evidence Centred Design* outlines a chain of reasoning that starts with domain analysis and then moves on to the crucial stage of domain modelling in which the test construct or ability is articulated: what claims are we making about the test-taker, what evidence do we need to substantiate those claims, and what tasks we will need to elicit that evidence. Bachman and Palmer (1996) describe this domain as the Target Language Use (TLU) situation. Kane (2012) presents the link between domain of use – observation of an individual's performance on a particular task – and the decisions that are made about an individual's ability as a chain of inferences as illustrated in *Figure 1* below.

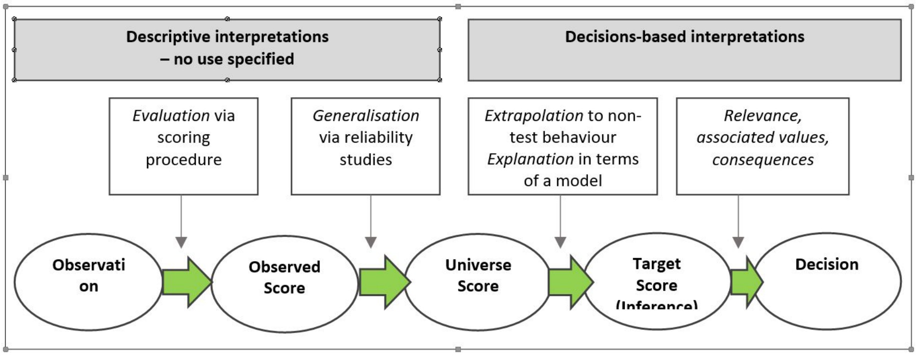


Figure 1: Kane's chain of inferences (McNamara & Roever, 2006)

In their operationalisation of what are relatively abstract validity theories, O'Sullivan and Weir (2011) connect the domain of use and the claim about an individual's ability through four questions focusing on test-taker characteristics (Who are we testing?), the construct or ability (what are we testing?), the tasks used to elicit that ability (How are we testing it?) and the assessment criteria (What system will be use to score it?); the interaction of these questions is presented in *Figure 2* below. They go on to explicitly state:

"Unless we can demonstrate empirically that… they demonstrate a link between the underlying concepts, our test is unlikely to allow us to make valid inferences." (2010, p. 23).

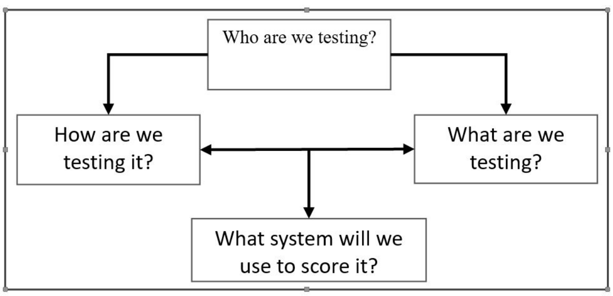


Figure 2: Operationalisation of validity theories (Weir& O'Sullivan, 2010)

What is echoed throughout these approaches and operationalisations of test validity is that the domain of use needs to be reflected in the tests. To put it in a different way: in order to make a plausible decision about whether someone has the ability to perform a certain communicative task in a certain communicative situation, the link between the test and the domain of use must be demonstrated. If a test is shown to misrepresent or underrepresent the domain to which it purports to link, then test validity is threatened.

Crucial to identifying whether a test is valid is understanding the intended domain of use or the TLU, in Bachman and Palmer's terms. One argument for continuing to use NS norms in testing is that the domain of use is characterised by standard forms of English (Berns, 2006 - see above). Is this indeed the case? In the following sections, I attempt to answer two questions:

- What does the domain / Target Language Use situation / underlying construct of English as a lingua franca in an EAP context look like?
- What argument can be developed to demonstrate the link between a test and the ELF construct and what evidence is needed to support this argument?

Finally, the argument structure is briefly applied to two tests by way of example.

3. The Domain of Use

English as a lingua franca

Definitions of ELF are as numerous as the different ways of referring to the broad concept of English, or Englishes, that are used as a common language of communication. The online Cambridge dictionary defines *lingua franca* as follows, using English in the example of use:

lingua franca

noun [C usually singular]

A language used for communication between groups of people who speak different languages: *The international business community sees English as a lingua franca.* (Cambridge Dictionary, n.d.)

Widdowson also refers to this commonality of communicative form between speakers who share no other:

"English as a lingua franca is the communicative use of linguistic resources, by native and non-native speakers, when no other shared means of communication are available or appropriate." (2013, p.190)

As does Seidlhofer: ELF is "communication in English between speakers with different first languages" (Seidlhofer, 2005, p. 339); later, she describes it as, "any use of English among speakers of different first languages for whom English is the communicative medium of choice, and often the only option" (Seidlhofer 2011, p. 7).

But to use only these definitions is to simplify the situation and to ignore the rapidly changing dynamics of global communication. Seidlhofer, in 2009, argued for a new perspective on ELF and particularly how it is being influenced by new technologies. Figure 3 below illustrates the traditional view of English as a global form of communication, the new way in which we could or should conceptualise ELF, and the catalysts driving this change.

The left-hand side of the diagram depicts the Circles that Kachru used to describe the Englishes and English use that are by this time, very familiar to most applied linguists. They include:

- the Inner Circle – varieties attributed to mother-tongue or 'Native Speakers', typically the UK, US, Australia, Canada, New Zealand and Ireland;
- the Outer Circle – describing varieties that have emerged following decolonisation of the British Empire in the 1950s and 1960s, usually nativised with a corresponding written form and strongly associated with identity in the post-colonial world, such as Indian English, Singaporean English, Nigerian English, Jamaican English; and,
- the Expanding Circle describing the use of English by those who do not fall into the first two Circles – a learned form of English, usually an Inner Circle form, generally assessed in relation to these standard forms, with deviations from these varieties described in terms of errors or fossilisation.

In Seidlhofer's 2009 argument for the re-evaluation of what defines English for the majority of users she points to two key 'push' factors:

- technology-enabled communication allows for increased international contact where we are no longer confined to communicating with people within our own immediate physical environments, but with a range of people from around the world;
- a move towards *communities of practice* rather than physical communities, academic communities being a good example, such as a language assessment and language learning community of practice.

The right-hand section of Figure 3 illustrates how global interaction and communities of practice cut across Kachru's circles, resulting in the need for a different conceptualisation of what constitutes ELF in different settings, a characteristic that Leung & Lewkowitz (2006) and Canagarajah (2007) have also pointed out.

Seidlhofer sums the situation up as follows:

"With the current proliferation of possibilities created by electronic means and unprecedented global mobility, changes in communications have accelerated and forced changes in the nature of communication. And for the time being anyway, it is English as a lingua franca that is the main means of wider communication for conducting transactions and interactions outside people's primary social spaces and speech communities. It seems inevitable that with radical technology-driven changes in society, our sense of what constitutes a legitimate community and a legitimate linguistic variety has to change, too." (2009, p. 238)

There have been moves towards a definition of the construct of ELF, with corpus linguistics driving much of the outcomes, e.g. the English as a lingua franca in Academic Settings corpus which draws on data from speakers of 51 different first languages, and the Vienna-Oxford International Corpus of English (VOICE) which "seeks to redress the balance [between the predominantly NNS of English and the NS-referenced linguistic description of the language] by providing a sizeable, computer-readable corpus of English as it is spoken by this non-native speaking majority of users in different contexts". The move away from seeing English as being an inherently NS-domain is also evident in the most recent CEFR review – the CEFR Companion Volume (2018) has removed all references to 'native speaker' in any of the can-do statements. Finally, Jenkins's proposal of a lingua franca Core for phonology is well-known but has been only minimally adopted, partly as a result of a paucity of supporting evidence. Isaacs cautions, "substantially more empirical evidence is needed before the lingua franca Core can be… adopted as a standard for assessment" (2013, p. 8).

The lack of construct definition for ELF and its fluid and dynamic character, in addition to socio-political factors, are possible reasons for the continued reliance on Inner Circle of English in high-stakes tests, even in the face of evidence that standard forms are reflected to a lesser degree than other varieties in the domain of use. For example, the International English Language Testing System (IELTS) refers to NS norms in both the Grammar and Vocabulary and Pronunciation evaluation criteria (my emphasis in bold):

G&V: produces consistently accurate structures apart from 'slips' characteristic of **native speaker speech**

Pronunciation: is easy to understand throughout; **L1 accent** has minimal effect on intelligibility

https://www.ielts.org/-/media/pdfs/speaking-band-descriptors.ashx?la=en

Pearson, another major testing organisation, does so, too: "Pronunciation reflects the ability to produce consonants, vowels, and stress in a **native-like manner** in sentence context" (my emphasis in bold) (2011, p. 12).

The Educational Testing Service (ETS) ETS SpeechRater programme, an automated rating system, likewise appears to establish a NS benchmark by using a 'pronunciation dictionary', based on NS standards (with some alternative pronunciations) (Xi et al, 2008).

The continued use of predominantly NS varieties in major language tests calls into question the validity of the assessments if they are used to decide whether an individual is able to function in a context that is not shaped according to NS norms of English communication. Although the construct of English communicative ability is fluid and changing, the wider context of use – the community of practice – will ultimately shape the construct definition; some of these domains are stable enough to offer a more solid description of the ELF construct associated with them. This is what we turn to next.

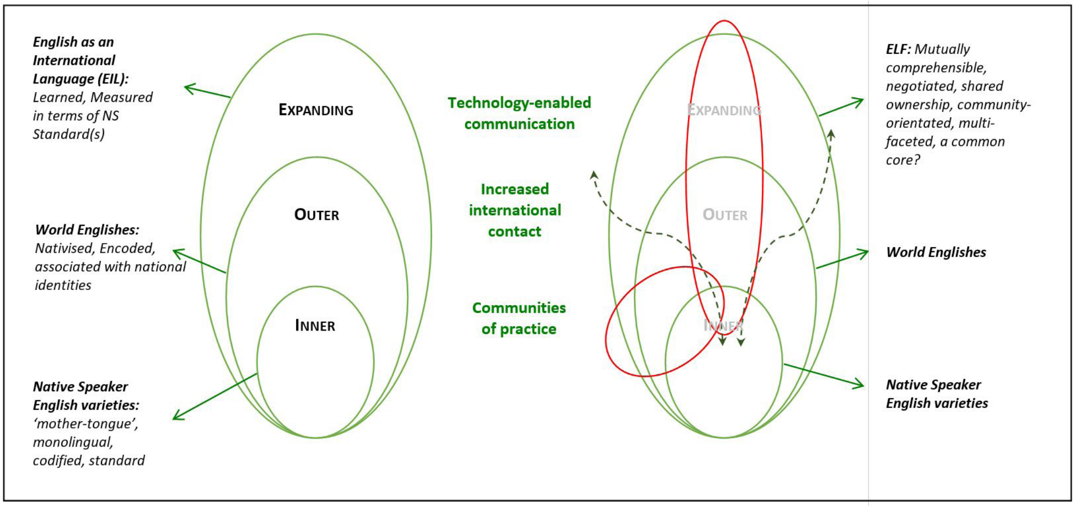


Figure 3: Changes in the conceptualisation of English(es) and English as a lingua franca

The domain of use

As suggested above, communities of practice are diverse, dynamic, and potentially overlapping. In the interests of brevity and the conciseness of an example validation argument, the scope of the discussion in this paper is limited to English for Academic Purposes (EAP), specifically the context of universities where English is the medium of instruction (EMI) and academic discourse. Given the large number of EMI institutions worldwide, the investigation of the domain of use is further limited to data for two major EMI destinations for international students, the UK and Australia. Indeed, EAP is one of the most prevalent uses of high-stakes, international tests of English proficiency. Before considering whether a test is valid as an instrument to decide whether someone's language ability is adequate to function in an EAP environment, the domain of use needs to be understood.

The following statistics allow us to better understand the EAP domain of use at universities in the UK and Australia. Figure 4 shows that more than two-thirds of the 2016-2017 cohort of students at UK universities were not from the UK but from a range of backgrounds, both EU and non-EU. Thus, it would follow that someone preparing for post-graduate study in the UK should expect to interact with a variety of fellow students from a wide range of language and cultural backgrounds, some NS, but the majority NNS.

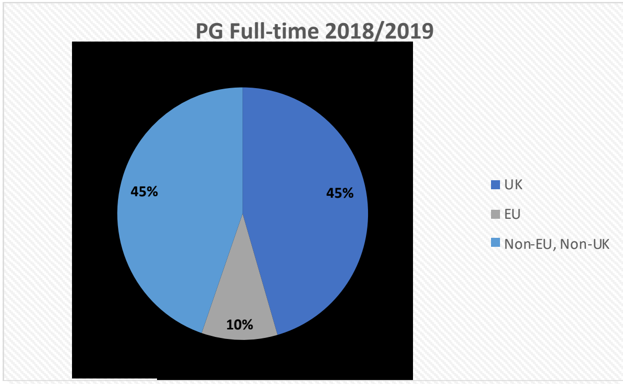


Figure 4: Higher Education student enrolments in Post-Graduate full-time study across the UK by domicile academic year 2018 - 2019.

Lecturers are also key stakeholders in the educational milieu of international students and they need to be understood and communicated with effectively. Universities UK, a collaboration of 137 universities across the UK puts the percentage of international staff working at UK universities at 30% (Universities UK international, 2018); Figure 5 shows that more than 30% of UK academic staff in 2018/2019 were not from the UK suggesting that, aside from regional UK accents, international students would need to understand and interact effectively with a range of lecturers and tutors from different backgrounds and with a variety of accents.

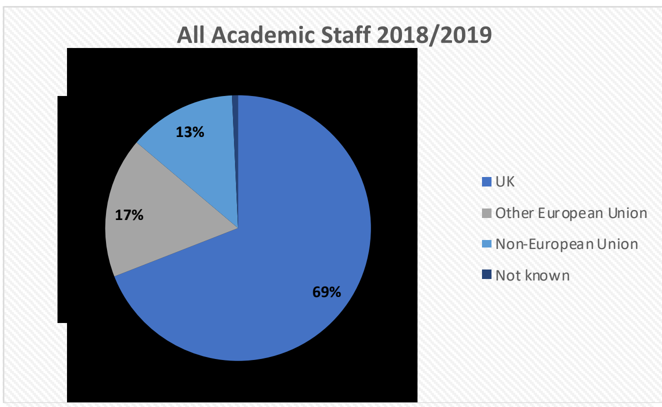


Figure 5: Percentage of academic staff employed at UK Higher Education institutions in the UK.

Figure 6 shows that the picture is no different in Australia where international enrolments are on the increase, meaning that students at university have a strong likelihood of interacting with someone who speaks a non-Inner Circle variety of English, or speaks English as their second or third language. At post-graduate level, international students make up over 40% of the study body (Figure 7).

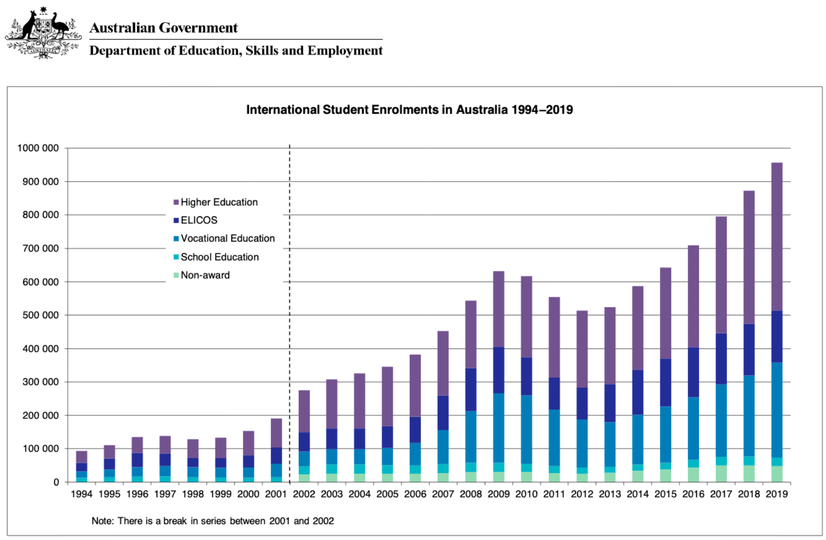


Figure 6: Rise in international student enrolments in Australia 1994 – 2019.

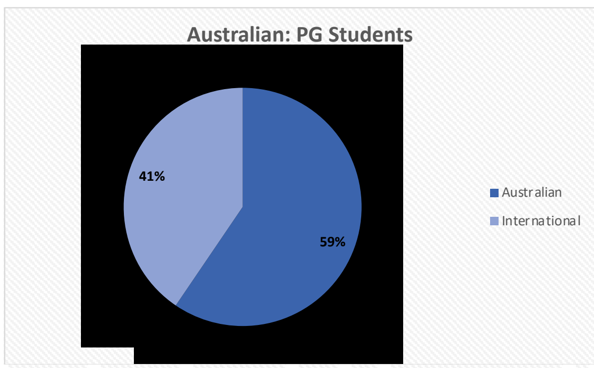


Figure 7: Percentage of on-campus post-graduate students enrolled in Australian Higher Education institutions by origin (possibly 2018 – exact year unclear from available data).

The recognition of this increasing internationalisation of Australian higher education is echoed in the press:

“Because the Government has effectively capped the number of domestic students, international students are becoming an increasing percentage of all students,” Mr Norton said.[1]

It is also investigated by research institutes: for example, the Grattan Institute reports that in 2018, just under three-quarters of students enrolled in Australian higher education institutions were Australian citizens or permanent residents.

Even this very superficial consideration of the domain of use – tertiary EMI institutions in two traditionally Inner Circle countries – suggests that NS norms and standard forms of English should not be the only varieties to be tested if linguistic preparedness for these contexts is the primary ability being evaluated. Instead, the context of use suggests that we should be evaluating someone’s ability

[1] https://www.abc.net.au/news/2018-04-18/australia-hosting-unprecedented-numbers-international-students/9669030

to communicate effectively with a range of speakers from different L1 backgrounds and that, rather than assessing proficiency in relation to a NS norm, test developers should be considering comprehensibility and ensuring that assessment for the situations described above is inclusive of variety as long as the comprehensibility principle is met.

4. A validity argument

The argument framework

Toulmin's argument schema is a tool for the evaluation of a claim (2003). Conceptualised as framework for the analysis of legal arguments, the schema is useful to guide test-developers in evidence-centred design and to support the validity assertions of their tests. It also provides a useful tool for those analysing the veracity of the validity claims of a language test. The latter is what is especially appealing about the framework in terms of language test evaluation: it helps to identify the types of evidence necessary to support a claim. An example of Toulmin's argument structure as applied in general to tests of spoken English within the ELF context is presented in Figure 8. Note that, in the interests of brevity and as the aim of this paper is to present an *example* of how the argument structure can be applied to the assessment of ELF, the details in Figure 8 pertain only to the evaluation of *spoken performances*. In the following section, an example of how this framework can be applied to evaluating the validity of two well-known English language is presented.

At the core of the argument are the facts necessary to support the overall claim of validity. In the case of ELF, we would want to know that the *Scores on a test of spoken English reflect a speaker's* ***ability to make themselves understood in an international context*** such as the EAP domains considered above. The facts (the grounds) that would act as the basis for this claim are that *the means of performance elicitation and the assessment criteria applied to the performances focus on* ***comprehensibility of and by a range of different L1 speakers.***

It is, of course, necessary to substantiate the facts in order to link the grounds to the overall claim. In our example, this link between claim and underlying grounds exists as long as *the observed performances in the test provide observed scores reflective of an ability to be* ***comprehensible to a wide range of English speakers from different L1 backgrounds*** – the warrant. To evaluate whether this condition has been met, concrete evidence needs to be presented. Figure 8 suggests three key areas in which evidence can be presented and according to which we might consider in evaluating the validity of the claim: the task types that are used to elicit the performance (are they reflective of the context of use? – assumption 1);the rating criteria (do they have comprehensibility rather than native-speakerness as a benchmark? – assumption 2) and, a question related to reliability – are the evaluation criteria applied accurately and consistently? – assumption 3). The data that could be used to substantiate or refute these assumptions are described in Figure 8, both qualitative as well as quantitative data that is necessary to support the overall claim.

Finally, legal-orientated rebuttals presented on the right-hand side of the diagram provide useful jumping-off points for a critical analysis of a language test being used to assess a test-taker's readiness to function in an ELF context:

- are the tasks on the test comprised of only NS linguistic and cultural input?
- is preference for a NS accent evident in the rating scales?
- who are the raters? are they made up only of NSs or is there adequate representation of a

range of proficient users of ELF? This applies to tests with human raters as well as tests where machines 'learn' from a pool of human raters.

Figure 8 presents just one example of how Toulmin's argument structure can be used to evaluate tests from the ELF perspective.

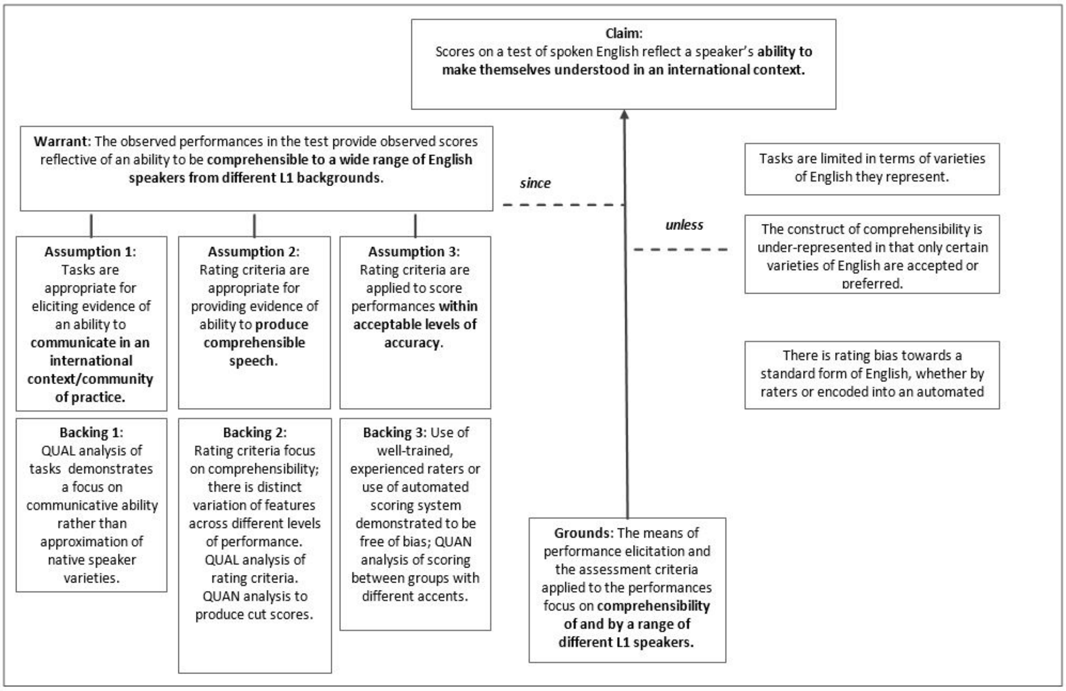


Figure 8: An argument structure for the evaluation of a language test in an ELF context

Applying the argument framework: an example

This section presents a brief analysis of two tests to illustrate the potential application of the argument framework presented above. This is by no means intended to be a detailed analysis of any of the tests; rather, the aim is to provide an example of how the strength of the link between test and domain of use can be investigated using the argument framework to establish the validity of a test.

Assumptions:

1. The tests are being taken as predictors of ability to communicate in the domain of use explored above, i.e. EAP in the UK/Australia.

2. The ability under scrutiny is the production of spoken English and, as such, the focus of the mini-analysis are the speaking modules or components thereof.

The two tests under consideration are the International English Language Testing System (IELTS) and the Pearson Test of English (PTE). These are both high-stakes tests, frequently required for entrance to higher education in the UK or Australia. The tests are different in that IELTS is delivered and rated by humans while PTE is delivered by computer and rating is automated, i.e. Artificial Intelligence (AI) is used to assign scores to spoken performances. Given that this is an illustration of the application of the argument framework above, one specific task was focused on for PTE due to the variety of tasks included in the integrated listening-speaking module; IELTS, however, is somewhat more homogenous in nature as the entire speaking test consists of a 12-14 minute Oral Proficiency Interview (OPI) with a trained interlocutor.

Table 1 below illustrates how the assumptions taken from figure 8 can be supported and rebutted for each of the two tests. While this illustrative analysis has drawn only on publicly available materials, a more robust analysis by the test developers or those engaged in evidence-based, critical selection of tests for EAP purposes could – and should – include independent research to obtain the necessary evidence.

The table below also shows that some evidence can support the validity of a test while other evidence undermines those claims. It also illustrates how an analysis of test validity must be linked to the purpose for which the test will be used.

Table 1: Application of the argument framework to two language tests – an example

IELTS [Overall speaking module]		PTE [Task 1.5 – Listen and Retell]	
For related citations and sources please see below.			
Warrant: The observed performances in the test provide observed scores reflective of an ability to be comprehensible to a wide range of English speakers from different L1 backgrounds.			
Supporting evidence	Rebuttal	Supporting evidence	Rebuttal
Assumption 1: Tasks are appropriate for eliciting evidence of an ability to communicate in an international context/ community of practice.			
Interaction with an interlocutor, providing a reasonably authentic communication context.	If all interlocutors are NS, the test context is not reflective of an international context and the construct is narrowed.	Task 1.5 is an integrative task that reflects the EAP setting, i.e. listening to a lecture and then speaking to summarise what was heard.	If the listening is always a NS, this does not reflect an international context and the construct is narrowed.
Assumption 2: Rating criteria are appropriate for providing evidence of ability to produce comprehensible speech.			
The analytical rating scales include reference to “intelligibility”. Academic research shows a link between scores and performance in an EAP setting.	The analytical rating scales include reference to “‘slips’ characteristic of native speaker speech”.	Scoring criteria include “how accurately and thoroughly” meaning is conveyed, i.e. a focus on content.	The scoring description includes reference to “regular speakers” of English; score guide explicitly lists “native-like” as the highest level of proficiency, above “advanced” for both fluency and pronunciation.
Assumption 3: Rating criteria are applied to score performances within acceptable levels of accuracy.			
Several means are in place to ensure consistency in the marking of the writing and speaking tests including robust recruitment and training, standardisation and monitoring, as well as statistical analysis of results.	Standard Error Measurement in human rating of productive skills may allow more tolerance of bias by some raters.	AI scoring removes potential human bias towards different accents, for example.	Educated, proficient or NS speakers score poorly on the test, e.g. reference below.
Sources for above			

IELTS public band descriptors (https://www.ielts.org/): *Produces consistently accurate structures apart from 'slips' characteristic of native speaker speech. [Grammar and vocabulary criterion]* *Is effortless to understand [Pronunciation criterion]* https://www.ielts.org/teaching/examiner-recruitment-and-training https://www.ielts.org/about-the-test/ensuring-quality-and-fairness *"The clearest finding emerging from this research is the predictive validity of IELTS scores in relation to general language performance."* Ingram & Bayliss: 2007, p. 59	PTE: Listen and Retell https://pearsonpte.com/the-test/format/english-speaking-writing/re-tell-lecture/ *Content is scored by determining how accurately and thoroughly you convey the situation, characters, aspects, actions and developments presented in the lecture.* *Pronunciation: Does your response demonstrate your ability to produce speech sounds in a similar way to most regular speakers of the language?* *Pronunciation is scored by determining if your speech is easily understandable to most regular speakers of the language. The best responses contain vowels and consonants pronounced in a native-like way, and stress words and phrases correctly. Responses should also be immediately understandable to a regular speaker of the language.* *PTE Academic recognizes regional and national varieties of English pronunciation to the degree that they are understandable to most regular speakers of the language.* PTE Score Guide for pronunciation: *5 Native-like 4 Advanced 3 Good 2 Intermediate 1 Intrusive 0 Non-English* News article (see references): *Irish vet fails oral English test*

5. Next steps

ELF is centred around the concept of mutual comprehensibility. In order to move towards testing of the comprehensibility of spoken English rather than an approximation to NS varieties, several steps need to be taken:

- achieving a better understanding of what constitutes comprehensibility;
- ensuring that research into comprehensibility is not limited to NS assessments of what is comprehensible but includes the perceptions of a range of speakers of English to reflect the real-world communication context;
- encouraging test developers to take active steps to better reflect the ELF context; for example, ensuring that listening tests include a range of accents, and removing reference to 'native-like' speech in rating rubrics;
- guarding against encoding bias into the algorithms of automated assessment systems by not relying only on NS reference points;
- raising awareness amongst test users that 'English' does not only include the Englishes of the Inner Circle and that the goal is to be comprehensible to listeners from a wide range of different language backgrounds; as consumers of commercial language tests, test users have the power to influence the test developers.

6. Conclusion

This paper presented the notion that real-life use of English is not limited to the use of standard forms of English but includes the use of a myriad of Englishes that facilitate common understanding: ELF. The construct definition of such varied, dynamic use of language presents challenges to the language assessment community but does not mean that it can be ignored – the consequences to test-takers and society more broadly are too great. While researchers work on understanding more about the underlying construct of ELF – a crucial component to more reflective, equitable and fit-for-purpose testing – critical questions should be asked about the validity of current language tests to drive test developers in the direction of a more equitable, fair and inclusive evaluation of this global lingua franca.

The question of whether a test is valid is inextricable from its purpose and the context of language use to which the test scores are linked. Where the communicative environment is peopled by so many different voices from different language backgrounds interacting in English, it is becoming increasingly necessary for language assessment tools to reflect this, and that test users – educational institutes, Ministries of Education, immigration agencies, employers and test-takers themselves – seek the relevant assurance that tests do, indeed, do this. The example of the argument framework presented in this paper demonstrates a powerful tool with which to identify the crucial questions that need to be asked and types of evidence that should be demanded as proof that a test is valid in a real-life context. While Toulmin's framework serves as a robust tool for critically evaluating tests in the context of ELF use, it can also provide a blueprint for test developers keen to be fair and inclusive in the design of their assessments to better reflect the real-life use of English.

References

Ammon, U. (2010). World languages: Trends and futures. In Coupland, N. (Ed.), *The handbook of language and globalization* (pp. 101 – 122). Oxford: Blackwell Publishing.

Australian Associated Press. (2017, August 8). *Computer says no: Irish vet fails oral English test needed to stay in Australia*. https://www.theguardian.com/australia-news/2017/aug/08/computer-says-no-irish-vet-fails-oral-english-test-needed-to-stay-in-australia

Australian Education Network, (2020). University Rankings. https://www.universityrankings.com.au/international-student-numbers/

Australian Government Department of Education, Skills and Employment. (2020). *International student enrolments in Australia 1994-2019*. https://internationaleducation.gov.au/research/International-StudentData/Pages/InternationalStudentData2019.aspx# Annual_Series

Bachman, L.F. and Palmer, A. S. (1996). *Language testing in practice*. Oxford: Oxford University Press.

Berns, M. (2006). Word Englishes and communicative competence. In B.B. Kachru, Y. Kachru, & C.L. Nelson (Eds.), *The handbook of world Englishes*, (pp 718 – 730).

Cambridge Dictionary. (n.d.). Lingua franca. In *Dictionary.Cambridge.org*. Retrieved 19 May 2020, from https://dictionary.cambridge.org/dictionary/english/lingua-franca

Canagarajah, S. (2007). Lingua franca English, multilingual communities, and language acquisition. *The Modern Language Journal, 91*(1), 923–939. doi-org.ezproxy.jyu.fi/10.1111/j.1540-4781.2007.00678.x

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

Council of Europe (2018). *Common European Framework of Reference for Languages: Learning, Teaching, Assessment. Companion Volume*. https://rm.coe.int/cefr-companion-volume-with-new-descriptors-2018/1680787989

Crystal, D. 2003. *English as a global language (Second Edition)*. Cambridge: Cambridge University Press.

Davies, A. (1999). Standard English: Discordant voices. *World Englishes*, *18*(2), 171–186. doi.org/10.1111/1467-971X.00132

Elder, C. & Harding, L. (2008). Language testing and English as an international language: Constraints and contributions. *Australian Review of Applied Linguistics, 31*(3), 34.1-34.11. doi:10.2104/aral0834

ELFA 2008. *The Corpus of English as a lingua franca in Academic Settings*. Director: Anna Mauranen. http://www.helsinki.fi/elfa (accessed 26 April 2020).

Galloway, N. (2018). ELF and ELT teaching materials. In J. Jenkins, W. Baker, & M. Dewey (Eds.), *The Routledge handbook of English as a lingua franca* (pp. 468-480). Routledge.

Graddol, D. (2006). *English next*. London: British Council.

Harding, L., & McNamara, T. (2017). Language assessment: The challenge of ELF. In J. Jenkins, W. Baker & M. Dewey (Eds.), *The Routledge handbook of English as a lingua franca*, (pp. 570–582).

Higher Education Statistics Agency. (2019). Figure 6 - All staff (excluding atypical) by equality characteristics 2017/18. Ref. ID: SB253 Figure 6. https://www.hesa.ac.uk/data-and-analysis/sb253/figure-6

Higher Education Statistics Agency. (2020). Figure 7 - HE student enrolments by level of study, mode of study and domicile 2018/19. Ref. ID: SB255 Figure 7. https://www.hesa.ac.uk/data-and-analysis/students/where-from

Ingram, D. & Bayliss, A. (2007). *IELTS as a predictor of academic language performance, Part 1*. In: IELTS Research Reports, Volume 7, 2007. British Council and IELTS Australia Limited.

International English Language Testing System (IELTS). *IELTS Speaking: Band Descriptors (public version)*. British Council, IDP: IELTS Australia and Cambridge English Language Assessment. https://www.ielts.org/-/media/pdfs/speaking-band-descriptors.ashx?la=en

Isaacs, T. (2013). Assessing pronunciation. In Kunnan, A. J. (Ed.), *The Companion to Language Assessment Vol. I* (pp. 140–155). doi.org/10.1002/9781118411360.wbcla012

Jenkins, J. (2000). *The phonology of English as an international language*. Oxford: Oxford University Press.

Jenkins, J. (2006). The spread of EIL: a testing time for testers. *ELT Journal*, *60*(1), 42–50. doi.org/10.1093/elt/cci080

Kachru, B.B. (1985). Standards, codification and sociolinguistic realism: the English language in the outer circle. In Quirk, R. and Widdowson, H. (eds). *English in the World: Teaching and Learning the Language and Literatures.* Cambridge: Cambridge University Press, pp. 11–30.

Kane, M. (2012). Articulating a validity argument. In G. Fulcher & F. Davidson (Eds.), *The Routledge handbook of language testing* (pp. 34–47). Oxford and New York: Routledge.

Kirkpatrick, A. (2007). *World Englishes: Implications for international communication and English language teaching.* Cambridge: CUP.

Leung, C. and Lewkowicz, J. (2006). Expanding horizons and unresolved conundrums: Language testing and assessment. *TESOL Quarterly, 40*(1), 211–234. doi-org.ezproxy.jyu.fi/10.2307/40264517

Macmillan Education Limited. (n.d.). *Global* Coursebook. http://www.macmillanglobal.com/about/the-course

McNamara, T. (2006). Validity in language testing: The challenge of Sam Messick's legacy. *Language Assessment Quarterly*, *3*(1), 31–51. doi:10.1207/s15434311laq0301

McNamara, T. and Roever, C. (2006). *Language testing: The social dimension.* Oxford: Blackwell.

McNamara, T. (2014). Evolution or revolution? *Language Assessment Quarterly*, *11*(2), 226-232. DOI: 10.1080/15434303.2014.895830

Messick, S. (1989). Validity. In R.L. Linn (Ed.), *Educational measurement (3rd ed.)* (pp. 13-103). New York: Macmillan.

Mislevy, R. J., & Yin, C. (2012). Evidence-centred design in language testing. In G. Fulcher & F. Davidson (Eds.), *The Routledge Handbook of Language Testing*. Oxford and New York: Routledge.

Nelson, C. (2011). *Intelligibility in World Englishes: Theory and Application.* New York: Routledge.

Norton, A. & Cherastidtham, I. (2018). *Mapping Australian higher education 2018.* Grattan Institute. Grattan Institute Report No. 2018-11, September 2018, https://grattan.edu.au/wp-content/uploads/2018/09/907-Mapping-Australian-higher-education-2018.pdf

O'Sullivan, B., & Weir, C. J. (2011). Test development and validation. In B. O'Sullivan (Ed.), *Language testing: Theories and practices* (pp. 13-28). Basingstoke: Palgrave Macmillan.

Paul, L.P., Simons, G.F. and Fennig, C.D. (Eds.) 2020. *Ethnologue: Languages of the World, Eighteenth edition.* Dallas, TX: SIL International. www.ethnologue.com [accessed: May 2020].

Pearson. (2020). *PTE Academic Score Guide for test takers. Version 12 – April 2020.* Palo Alto: Pearson.

Peters, P. (1995). *The Cambridge Australian English style guide.* Melbourne: Cambridge University Press.

Quirk, R. (1985). The English language in a global context. In R. Quirk, R. & G. H. Widdowson (Eds.), *English in the world: Teaching and learning the languages and literatures* (pp. 1-6). Cambridge: Cambridge University Press and The British Council.

Robinson, N. (2018, April 18). Australia hosting unprecedented numbers of international students. *ABC News.* https://www.abc.net.au/news/2018-04-18/australia-hosting-unprecedented-numbers-international-students/9669030.

Seidlhofer, B. (2005). English as a lingua franca. *ELT Journal 59*(4), 339–341. doi.org/10.1093/elt/cci064

Seidlhofer, B. (2009). Common ground and different realities: World Englishes and English as a lingua franca. *World Englishes*, *28*(2), 236–245. doi.org/10.1111/j.1467-971X.2009.01592.x

Seidlhofer, B. (2011). *Understanding English as a lingua franca.* Oxford: OUP

Toulmin, S. (2003). *The uses of argument (Updated edition).* Cambridge: Cambridge University Press.

Universities UK International. (2018). *International facts and figures: Higher Education 2018.* https://www.universitiesuk.ac.uk/policy-and-analysis/reports/Documents/International/International%20Facts%20and%20Figures%202018_web.pdf

VOICE. 2013. *The Vienna-Oxford International Corpus of English* (version 2.0 XML). Director: Barbara Seidlhofer; Researchers: Angelika Breiteneder, Theresa Klimpfinger, Stefan Majewski, Ruth Osimk-Teasdale, Marie-Luise Pitzl, Michael Radeka. https://www.univie.ac.at/voice/ (accessed 26 April 2020).

Widdowson, H. G. (2013). ELF and EFL: What's the difference? Comments on Michael Swan. *Journal of English as a lingua franca*, *2*(1), 187-193. doi: https://doi-org.ezproxy.jyu.fi/10.1515/jelf-2013-0009

Williams, R. O. (1980) Our Dictionaries. In J. Green (Ed.), *Chasing the Sun: Dictionary-Makers and the Dictionaries they Make,* p. 379. London: Pimlico/Random House.

Xi, X., Higgins, D., Zechner, K., & Williamson, D. M. (2008). Automated scoring of spontaneous speech using SpeechRater[SM] V1.0. *ETS Research Report Series*, *2008*(2), i–102. doi.org/10.1002/j.2333-8504.2008.tb02148.x

ĐÁNH GIÁ KỸ NĂNG SỬ DỤNG TIẾNG ANH NHƯ MỘT NGÔN NGỮ TOÀN CẦU: LẬP LUẬN VỀ TÍNH GIÁ TRỊ

Sheryl Cooke

Hội đồng Anh, Đại học Jyväskylä, Phần Lan
989 Đường Tây Bắc Kinh, Thượng Hải, 200041, Trung Quốc

Tóm tắt: Việc sử dụng tiếng Anh trong thực tiễn thường có sự tham gia của những người có nền tảng ngôn ngữ khác nhau với mục tiêu chính là sự hiểu nhau, và phần lớn các cuộc hội thoại bằng tiếng Anh này không có sự tham gia của những người nói tiếng Anh bản ngữ (Graddol, 2006; Kirkpatrick, 2007). Tuy nhiên, thay vì tập trung vào tính dễ hiểu của lời nói, nhiều bài kiểm tra vẫn đo lường khả năng nói của thí sinh với tham chiếu về một khuôn mẫu bản ngữ lý tưởng. Điều này khiến cho tính giá trị của bài thi bị giảm trong việc đánh giá khả năng giao tiếp nói khi tiếng Anh được dùng như một ngôn ngữ toàn cầu, dẫn tới việc bỏ sót kỹ năng cần đánh giá hoặc đánh giá các yếu tố không liên quan.

Việc xác trị một bài thi tiếng Anh như một công cụ để giao tiếp toàn cầu bao gồm chứng minh mối liên hệ giữa kỹ năng cần đánh giá (khả năng giao tiếp thực tiễn trong trong bối cảnh cụ thể) với các tác vụ trong bài thi và tiêu chí đánh giá (McNamara, 2006). Các bằng chứng hỗ trợ việc giải thích ý nghĩa của điểm số cần được trình bày như một phần của lập luận tổng thể về tính giá trị. Trước hết, bài viết này muốn chỉ ra bối cảnh sử dụng tiếng Anh mà nhiều thí sinh trong các kỳ thi lớn hướng tới với mục đích học thuật thường là bối cảnh trong đó tiếng Anh được dùng như một ngôn ngữ toàn cầu (English as a lingua franca – ELF). Tiếp theo, lập luận của Toulmin (2003) được tận dụng để tìm ra những bằng chứng cần thiết cho việc chứng minh các khẳng định về một bài kiểm tra có tính đại diện tốt cho khả năng sử dụng tiếng Anh như ngôn ngữ toàn cầu. Mô hình được đưa ra và áp dụng vào phân tích minh họa hai bài thi Nói tiếng Anh. Bài viết có mục đích đưa ra minh chứng về mặt thuyết cho sự cần thiết của việc tập trung vào tính dễ hiểu trong giao tiếp và việc tăng cường sự đa dạng ngôn ngữ trong đánh giá ELF. Ngoài ra, tác giả cũng mong muốn đưa ra một mô hình xác trị có tính ứng dụng đối với người soạn và sử dụng bài thi.

Từ khóa: Tiếng Anh như một ngôn ngữ toàn cầu, tính giá trị của bài thi, tính dễ hiểu

ENHANCING STUDENTS' PROBLEM IDENTIFICATION IN TRANSLATION TEACHING IN AN ENGLISH LANGUAGE PROGRAM

Nguyen Thi Thu Huong*

University of Foreign Language Studies, University of Danang, 131 Luong Nhu Hoc, Cam Le, Danang, Vietnam

Received 16 April 2020
Revised 07 June 2020; Accepted 22 July 2020

Abstract: Teaching translation in Vietnam and other contexts tends to overemphasise linguistic issues in translation and lack focus on students' translation process. Drawing on a functional approach to translation, this pedagogical study investigates students' identification of translation problems (part of strategic competence) in a translation course in a tertiary English language program. The differences in students' ability to identify translation problems were evaluated through a contrastive examination of a qualitative analysis of students' written reflections on the translation tasks and their translation solutions both before and after the workshops. The study found that students diverted their attention from linguistic problems to other non-linguistic ones (extralinguistic issues related to extratextual features and general style conventions) after the workshops. This change led to some students' attempts to avoid word-for-word translation and produce more idiomatic translation solutions as shown in the analysis of titles. Recommendations were made on the inclusion of text analysis and problem identification in developing students' translation skills.

Keywords: translation teaching, translation process, translation problems, translation problem identification, text analysis

1. Introduction

The pressures caused by the increased global and local demand for translation services and the lack of professional translator training in Vietnam are directed to translation courses in foreign language programs at the tertiary level which are expected to train graduates with adequate skills for employment. Nevertheless, graduates of English programs in Vietnam who are often expected to engage in professional translation work usually lack the skills and ability to do so (Ho & Bui, 2013; T. T. Nguyen, 2010). This can be linked with the problematic translation teaching at the tertiary level in Vietnam, which is usually part of language programs. Some studies (Ho and Bui, 2013; T.T. Nguyen, 2010; Pham and Ton, 2007; Pham and Tran, 2013) reported that students' translation ability was limited due to their lack of an adequate understanding of the source text and their difficulties in finding target language equivalents. Students also complained that they were not equipped with translation strategies and analytical

* Tel.: 84-787582425
Email: huong2kh@gmail.com; ntthhuong@ufl.udn.vn

This research is funded by Funds for Science and Technology Development of the University of Danang under project number B2019-DN05-25

ability in dealing with a translation task. Ho and Bui (2013) found that the most common problems encountered by students are related to word usage and grammar errors. In other words, students mainly focus on linguistic aspects of translation and linguistic problems and in many cases and they tend to produce too literal or word-for-word translations.

Other studies done in many countries including Indonesia, Iran, China and Thailand (Avval, 2013; Chen, 2010; Saleh & Weda, 2018; Wongranu, 2017) reported similar problems in translation teaching. Saleh and Weda (2018) who investigated the poetry translation of students in an English program in an Indonesian university revealed that many students had problems and difficulties in translating vocabulary and sentence structures. Wongranu (2017) found that students in Thailand made grammatical errors due to their use of read and translate procedures and they produced too literal translation. This led to students' anxiety and lack of confidence in translating. In these classes, teachers hadthe tendency to draw students' attention to the source text (ST) and its linguistic translation problems. One possible explanation for the overemphasis of linguistic issues and the resulting impact on students' translation is that both educators tend to be more interested in linguistic differences than students' analytical skills and strategies in translation.

While translation is still considered a purely linguistic activity in several contexts, many researchers (Carreres & Noriega-Sánchez, 2011; Cook, 2010; Leonardi & Salvi, 2016) in their efforts to improve students' translation agree that translation is a communicative act that refers to the use of translation in a particular situation. They maintain that instead of overemphasising linguistic issues and students' linguistic capacity in translation, educators should pay attention to other translation issues including problems with text types or translation situations. This study adopts the view of the functional approach to translation which goes beyond the linguistic issues in raising students' awareness of different translation problems and improving their translations. The study will inform translation teaching at the tertiary level in Vietnam and other countries.

2. Literature review

This section presents problem identification and justification in teaching translation. It will also present key elements of the functional approach to translation that will be beneficial to enhancing students' ability to identify problems and justify their translations.

2.1 Problem identification and problem-solving

Problem identification and justification are considered aspects of translator's strategic subcompetence, a component of translator competence by a translation research group called PACTE (2008) (Fernández and Zabalbeascoa, 2012a, b). Strategic subcompetence "creates links between the different subcompetences (eg., bilingual, extra-linguisitic, instrumental and psycho-physiological competentences) as it controls the translation process" (PACTE, 2008, p.107). By activating the other subcompetences, strategic subcompetence maintains the efficiency of the translation process by enabling translators to identify translation problems and apply procedures and methods to solve the problems.

Although PACTE's definition and strategic subcompetences are specific to professional translator training, it is relevant to translation teaching at tertiary level where students' strategic awareness of the translation process is crucial. Many process-oriented pedagogical proposals including those by Kokkinidou and Spanou (2013), Lee and Gyogi (2018) and Leonardi (2010) have attended to developing students' translation process as part of their translation skill development. Process-oriented pedagogical approaches allow learners to be conscious of varied factors in the translation process and make informed translation decisions.

In this study, the translation process derives from Gile's (2009) model of translation that includes: comprehension of the ST and reformulation or production of the target text (TT). Gile explains that the actual process involves the translator reading the whole ST to identify problems and their attempting to solve them; therefore, the translation process can refer to problem identifying and problem-solving. It is possible to propose a definition of translation problem at this stage. A translation problem refers to a text segment (verbal or non-verbal) that is either at micro level (i.e., a text segment) or at macro level (i.e., at the text level) and that requires the translator to consciously apply a justified translation strategy (González-Davies & Scott-Tennent, 2005). This definition indicates the interrelated counterparts of problems and strategies. In translation, problems and difficulties can be identified in both phases: comprehending the ST and producing the TT. The problems and difficulties are then solved by translation strategies. This study will look into students' reporting of their problems and difficulties in translation and their solutions or strategies to some of the ST items. The next section introduces elements of the functional approach to translation in promoting students' problem identification.

2.2 A functional approach to translation: text analysis, translation briefs and types of problems

The functional approach to translation emphasises "skopos", or function which determines how a translated text is produced (Nord 1997). Translation must be fit or adequate for purpose. A text may be translated in several ways depending on the TT purpose which can be assigned by the initiator (a person who requires the translation). This approach has marked a transition from the linguistic view of translation (i.e., translation is considered a linguistic contrastive analysis activity) to a communicative intercultural activity (Schäffner & Wiesemann, 2001).

Nord's model has its practical use in translator training as students' competence in translation can be developed by taking into account the three aspects: ST analysis, the translation brief, and the hierarchy of translation problems (Nord 1997). Text analysis refers to the analysis of both extratextual and intratextual factors of the ST and TT. Extratextual factors include "sender" (text producer or writer), "sender's intention", "audience" (reader), "medium" (channel), "place of communication", "time of communication", "motive for communication" (why a text is produced), and "text function". Intratextual factors include subject matter, content, presupposition, text composition (or structure), non-verbal elements, lexis, sentence structure and suprasegmental features (e.g., italic or bold type). While ST analysis promotes translator's understanding of the ST, *translation brief* enables the translator to establish why a translation is required and by whom, what the clients need, and when, where the TT will be used, and who the TT addressees are.

As of translation problems, Nord's

translation problems are pragmatic (related to differences in the situations of the ST and TT), cultural translation problems (related to cultural differences), linguistic (related to differences between languages), and text-specific (e.g. metaphors or puns). Nord (1997, 2005) advocates that the translator should, in the first place, consider pragmatic perspectives in doing translation, giving priority to problems arising from the situations of the ST and the TT and the function of a translation. Nord differentiates between translation difficulties and translation problems. While difficulties are subjective and dependent on each translator, problems are more objective and seem to exist regardless of the level of translators.

A number of studies have proposed the incorporation of functionalist insights, particularly those of Nord into translation teaching in language programs. Colina and Lafford (2018) illustrate examples of translation activities that focus on the effects of contextual features (e.g., text, author, reader, and function) on understanding and producing texts. They include authentic texts and translation briefs so that students can understand how authentic texts are constructed in various genres, fields and contexts, keeping in mind different purposes and readers. Specifically, to assist students in their translation processes, the authors introduce both top-down and bottom-up genre-based approaches to text analysis where different elements of texts are considered.

Károly (2014) adopts a functional approach in supporting the enhanced status for translation in English language teaching. The author implemented the functional theoretical framework into foreign language programs with the aim of developing students' language and translation competence in an undergraduate English program in Hungary. After text analysis activities based on Nord's (2005) functional text analysis model were introduced in teacher-student, group and class discussions, the students had one week to translate three texts of different genres with translation briefs. The students' translations and discussions of the TT were analysed to identify and explore students' translation problems and difficulties (based on Nord's category of problems). The study found that linguistic problems were common among the students while they could identify other pragmatic and convention-related problems. The study illustrated the effective use of a functional translation approach in terms of students' identification of Nord's translation problems and difficulties. However, the study was implemented on a small scale, and it was mainly focused on data on students' errors that resulted from students' inability to deal with translation problems. To have a better insight into the process of problem identification, further research on students' reflections on their processes is needed.

Chen's (2010) study demonstrates the feasibility of incorporating functional approaches into translation teaching in general and Nord's model in particular in English programs. Text analysis, particularly consideration of text type and text function enables students to make any changes that fulfil communicative function(s) of a translation. The study particularly focused on the problem-solving process in which students identified problems in translating a ST item (i.e., metaphors) and adopted strategies to deal with them. The study reported a considerable change in students' treatments of metaphors. A student, for example chose to omit the metaphor in the translation due to a change in the function of the translation. Despite the strengths, the research had some methodological flaws. The same texts and examples of metaphors were used in the experimental teaching and the post translation task, which questioned the validity and reliability of the results. Analysis of the students' translations into English could have been affected due to their lack of competence in English. The present study presents qualitative data on the changes in students' translation process while translating into Vietnamese L1 before and after the experimental teaching in a series of workshops.

The study aims to answer the following questions:

1. What translation problems and difficulties do the students notice before and after the workshops?
2. How are students' translations different before and after the workshops?

3. Methods

3.1 Participants

Thirty students who were taking the introductory translation course (Translation 1) in a Vietnamese foreign language university were recruited to participate in this study. The participants had completed their first two years of skill training in English and other areas (e.g., computer skills). The students had taken compulsory courses in integrated English skills which aimed to help them consolidate level B1 English proficiency after the first year and achieve level B2 after the second year. Some written errors are still common among B2 students. During their first two years, the students had not undertaken any courses in translation.

3.2 The research procedure

Firstly, 30 third-year students performed Translation Task 1 in which they translated two texts and wrote about the difficulties they encountered during their translation processes. The second procedure involved the participation of the 30 students in three consciousness raising (CR) workshops. Lastly, all of the 30 students completed Translation Task 2, which was similar to Translation Task 1.

3.3 Translation workshops

There were three 150-minute workshops, each of which consisted of three 50-minute sections. Workshop 1 drew the students' attention to the notion of a good translation through discussions of their previous assumptions about translation, and their general ideas about their types of text encountered. Workshop 2 focused on text analysis which include different elements of texts and translation briefs or translation instructions. Specifically, the students were asked to translate one passage with and without translation briefs and they were then asked to identify differences in their translations. The aim of this activity is to develop the students' awareness that the translation brief enables the translator to make informed decisions during the decision-making process. I did not include types of translation problems in order to avoid bias during the study. In Workshop 3, the students were asked to identify the used strategies in the published Vietnamese translations and discussed the appropriateness of the strategies. This would lead to the students' awareness that the choice of strategy is based on text features, and the translation brief. In the final part of the workshop, the students were encouraged to write about what they considered to be accurate or what they thought constituted a good translation.

3.4 Translation tasks

Before and after the translation workshops, students were asked to do similar translation tasks. Each translation task consisted of two smaller tasks (Text 1a and Text 1b–carried out before the workshops; Text 2a and Text 2b–carried out after the workshops). Each task sheet featured a section for the students' translation, followed by a section for their comments about the problems and difficulties they encountered while translating (Writ. 1–comments before the workshops; Writ. 2–comments after the workshops). The translation tasks were similar in terms of structure, reference tools, time allocation and text length, translation to first language (L1) direction, text types, topics, text complexity, text features, students' flexibility with text sequence, and translation briefs (or instructions). The translation task design and the choice of texts, on the one hand, was appropriate to students' English proficiency to facilitate their comprehension and enable them to perform well within their capacity. On the other hand, the above-mentioned considerations allow for an optimal investigation of students' awareness of the translation process.

3.5. Written comments

After translating, the students were asked to write at least five of the problems and difficulties they had in doing the translation. The students' written comments followed the principles of the Integrated Problem and Decision Report (IPDR) developed by Gile (2004). The IPDR is a tool for studying the students' decision-making process in translation. Students are asked to note their problems and decisions in the translation task. According to Gile, students' notes can reflect their thoughts during the translation process and provide "information about students' problems, both individual and collective, and information about their translation strategies" (Gile, 2004, p. 2). The data analysis included students' comments in English which contained some inaccuracy due to their limited English proficiency.

4. Results and discussion

This section presents the results about students' reflections about their problems and difficulties before and after translation workshops. The analysis of students' translations to some ST items is also included to demonstrate relations between problem identification and translation solutions.

4.1 Students' problems and difficulties

Table 1. Total numbers of translation problems and difficulties

	Types of problems	Written comment 1 (before the workshops)	Written comment 2 (after the workshops)
1	Difficulties related to extratextual features		
	Author	3	8
	Reader	2	19
	Text function	1	5
	Subtotal	6	32
2	Difficulties related to general style conventions	6	16

3	Difficulties related to vocabulary and expressions		
	ST comprehension	34	20
	TT production*	28	25
	Both ST comprehension and TT production	2	1
	Unspecified difficulty**	18	24
	Sub-total	82	70
4	Difficulties related to sentences and sentence structures		
	ST comprehension	3	2
	TT production*	9	6
	Both ST comprehension and TT production	0	1
	Unspecified difficulty**	1	9
	Sub-total	13	18
	Total	107	136

* *TT production*: the difficulty in producing the provisional TT segment.

** *Unspecified difficulty*: The students copied the language in the ST and they did not specify whether the difficulty was related to ST comprehension or TT production. As there was not information about the students' problem, this category was not given a detailed analysis. Nevertheless, it still counted in the subtotal in each category.

In the analysis, the students' comments were classified into translation problems and difficulties related to extratextual features, general style conventions, vocabulary and expressions, and sentences and sentence structures (Table 1 above).

4.1.1 Difficulties related to extratextual features

The translation problems related to extratextual features referred to the students' comments about extratextual features including the author and text function. That the students' raised an issue was itself a sign of progress or improvement, notwithstanding the nature of their comments.

Before the workshops, very few comments referred to difficulties related to extratextual features. The students made only six comments about problems related to the author, the text function and the reader; however, the students did not elaborate on these problems. Three comments were concerned with the expression of the author's intentions (e.g., "I can't express all emotion which author want to convey" (Student 1, Writ. 1) or "Choose the meaning of words to be suitable for the intention of writer" (Student 19, Writ. 1). Two students referred to the text function and/or the reader. Student 28 was concerned about how target readers would comprehend their translation while briefly noting "the purpose of this text". Student 8's comment referred to text function by stating that the reader should be persuaded to do something (i.e., "carry out such as the target"). However, what Students 28 and 8 meant by text function was not clearly indicated.

Style: It is difficult to express the source text and persuade the reader carry out such as the target. (Student 8, Writ. 1)

After the workshops, the number of problems and difficulties was five time higher (32 versus 6). Firstly, eight comments referred to problems related to the author and understanding the author's intention (e.g., Students 4 and 10). Some other students (1, 18 and 21) also expressed difficulties in rendering the intentions, messages, and feelings of the author. Student 1 referred to the "feeling of author" in translating the "dead leaf blown in the wind" for which she used a word-for-word translation strategy. However, the reason for her dissatisfaction with her solution was not revealed.

Maybe I still don't express all feeling of author for "dead leaf blown in the wind." (Student 1, Writ. 2)

Secondly, after the workshops, some students indicated the difficulties they experienced in identifying and maintaining the purpose of the translation. Student 23 stated that they had difficulty defining the purpose of Text 2b ("I have problem when I have to define the type of text, purpose of text"). Student 1 referred to the function of the text ("persuasive") to the specific text type ("advertisement"—Text 2a—which is a government fact sheet, though). Similarly, Student 22 raised the problem of their translation's informative function ("my translation don't give information clearly"). Student 11 cited "easy green cleaning" in referring to their difficulty with "the goal of text". These students could have kept in mind the function or the purpose of translation in translating. However, as they did not elaborate much on their comments, it was hard to explore their understanding of text function, how it affected their translation solutions, and whether it played a role in their reflections on their translations.

Finally, in post-workshop written reflections, 19 comments (nearly two-thirds of the subtotal of the difficulties related to extratextual features) were devoted to the difficulties in determining the reader and translating for the reader. Compared with author and text function, the students elaborated further on this feature and they were able to see the importance of considering the reader in translating. For example, Student 21 maintained that, "You must define the reader (the receptor) and translate are accord". Only one comment pertained to the student's inability to determine who the reader was ("I don't know exactly who the readers are", Student 1, Writ. 2) while 14 comments clearly indicated who the translations targeted (e.g., children, Vietnamese children, or Vietnamese communities in Australia). In some cases, responses indicated that once the students were able to identify their reader, they seemed to be confident with their translation solutions, for example, to "green cleaning" in Title 2a (Text 2a)—"Easy Recyling—Create Your Own Eden" (Students 4 and 19) and "he" in Text 2b (Student 4).

Reader is Vietnamese communities in Australia, so I used exactly the word "green cleaning". (Student 4, Writ. 2)

After the workshops, students' reference to readers and/or other extratextual features suggested the impact of the workshops in which the students were instructed to consider the issue of readers and other extratextual features. Using the information in the translation brief they were given about the reader (i.e., Vietnamese communities in Australia), Students 4 and 19 justified their keeping the English ST "green cleaning" in their Vietnamese translations. Even though the solution did not match that of the published translation, this solution type was reasonably justified by the students.

The volume of comments on problems related to extratextual features after the workshops was five times the number raised in the corresponding evaluations completed prior to the workshops. The students' elaborations on the problems were limited, though. Obviously, the students' increased consideration of more problems with extratextual features might be linked with their exposure to text analysis activities in the workshops in which they were made aware of text features, particularly extratextual ones in translating.

4.1.2 Difficulties related to general style conventions

Difficulties with general style conventions were concerned with indicating the text types and conveying general stylistic features of the ST. This also covered the difficulties in deciding the formality of texts indicated by the choice of Vietnamese personal pronouns. The students' awareness of this type of issue seems to have allowed them to deal with problems at text level, particularly those related to style and/or choice of Vietnamese personal pronouns.

Before the workshops, six difficulties or problems related to general style conventions were raised. Four of the six problems mentioned were concerned with general differences in the writing style between English and Vietnamese authors (Student 6) or general assumptions about the text type (e.g., Students 3 indicated that the translation of the story should flow smoothly).

I have some difficulties when I must have the ability to approach and conversant with English texting style and then replace them into Vietnamese in the most suitable way to understand. (Student 6, Writ. 1)

Student 25 did not elaborate on their difficulty in translating "he" in Text 1b: "After reading the whole text, I think Marcus is young boy, so I translate "he" into "cậu bé" [a neutral Vietnamese personal pronoun for a boy]" (Student 25, Writ. 1).

While prior to the workshops, only 6 problems and difficulties were reported without much elaboration, after the workshops, the students listed 16 problems and difficulties with general style conventions, focusing on specific text types and formality in translating personal pronouns. Some students categorised the texts as advertisements (Text 2a) or funny stories (Text 2b). This reporting which was not found in the pre-workshop written comments seems to have influenced their translation approaches after the workshops. For example, Student 13 said that they had difficulty in expressing the ST as an advertisement. Students 5, 15, and 24 found it hard to translate the humour of the ST (Text 2b).

How to translate in a funniest way cos' this is a funny story (Student 5, Writ. 2)

Text form: I guess it is an advertising. Don't know how to express it as an advertising. (Student 13, Writ. 2)

I am not good at making the translation to be funny (because this is a funny story). (Student 15, Writ. 2)

The students did not elaborate much on the role of specific text type conventions, which may be due to the lack of focus on these conventions during the workshops. However, after the workshops, students made more comments about the choice of personal pronouns which were relevant to text formality. The use of Vietnamese personal pronouns is subject to the text type and its degree of formality. For instance, in a formal speech or text (e.g., an editorial or a political speech), the pronoun "you" in English is equivalent to "quý vị" that indicates a distance between writer/speaker and reader/listener. Meanwhile, in an informal speech (e.g., a letter to a friend), "you" has the meanings of "bạn" or "cậu" which suggests a degree of informality of the text type. Those students

(3, 4, 16, 19, 24, and 30) who referred to the problems in choosing the Vietnamese equivalents must have been aware of such differences in the use of personal pronouns in English and Vietnamese. They could have understood that English personal pronouns could be replaced with Vietnamese equivalents whose connotative meanings can be negative, positive, or neutral depending on the contexts of communication.

Use pronouns in Vietnamese, I use domestication when I translate but if in English the author used I – you, in Vietnamese, it may be "cậu – tôi", "nó – tôi". ["cậu – tôi" and "nó – tôi" are often used in informal situations but the latter sounds unfavourable] [Text 2b] (Student 19, Writ. 2)

Going one step further, some students explicitly discussed the appropriate choice of Vietnamese equivalents for English personal pronouns. A typical example was demonstrated by Student 30, who was conscious that the choice of certain Vietnamese words needed to take into account the content (i.e., the relationship of characters) in the story (Text 2b). They were, therefore, concerned about the connotations of the Vietnamese pronouns.

Translate pronouns, objects adjectives nouns into V with negative meanings, attitude. Example: … He/him: "nó", "hắn" or "anh ta" [For this student, these three Vietnamese equivalents were negative even though they may not be in other contexts]. (Student 30, Writ. 2)

After the workshops, the students listed problems with general style conventions three times more frequently than they had previously, with a greater focus on specific text types and formality in translating personal pronouns. They also provided more explanations for their difficulties with reference to specific text types, text formality, and content.

4.1.3 Difficulties related to vocabulary and expressions

The difficulties related to vocabulary and expressions referred to the students' comments about their comprehending and translating of vocabulary and expressions, such as idioms, metaphors, and similes. The problems they raised about vocabulary and expressions included:

- ST comprehension: The students had difficulty in understanding the meanings of vocabulary and expressions. They stated that they lacked vocabulary and did not know specific vocabulary items in the text.

- TT production: The students had difficulty in translating or finding Vietnamese translation words for vocabulary and expressions. The students wrote that even though they understood the vocabulary, they found it hard to translate.

- Both ST comprehension and TT production: The students mentioned both kinds of problems (comprehension and production) at the same time in one comment.

Before the workshops, the students mainly commented on linguistic problems and difficulties related to vocabulary and expressions (82 comments). Among these problems, those concerning the comprehension of vocabulary and expressions accounted for the largest portion, with 34 comprehension problems. Many students attributed their comprehension problems to their lack of knowledge of vocabulary and expressions. They noted that the words and phrases were "new", "strange," or "difficult" to them. Two students (19 and 29) attributed the comprehension problems to their lack of knowledge of the topic, while one student (25) said that they did not know what the pronoun "it" stood for in the sentence (this was concerned with grammatical knowledge).

Missing the knowledge of environment so I don't know some words such as disposed, sharps [Text 1a]. (Student 29, Writ. 1)

Subject "it"- "it was as silent as the grave". I don't know if "it" refers to the house or refers to the atmosphere of the house [Text 1b]. (Student 25, Writ. 1)

After the workshops, the number of problems in comprehending vocabulary and expressions decreased to 20 problems. Some students also referred to their insufficient knowledge of vocabulary and expression. One student found it difficult to understand the meaning of the ST's figurative language rather than the words themselves. This showed a certain change in their comments about the ST language. In other words, they began to take into account implied meaning(s) in addition to the surface meanings of words.

Misunderstanding some images: I was on my own; wander like a dead leaf blown in the wind. (Student 6, Writ. 2)

In TT production, before the workshops some students said that it was difficult to find the appropriate Vietnamese equivalents as there were many definitions of each word.

Some vocabulary have many definition and I can't choice suitable one. For example: Eden, sharp, syringe. (Student 1, Writ. 1)

Only a few students provided explanations for their problems. Some students found the type of vocabulary to be problematic. For example, Students 4 and 5 said that the titles were difficult to translate, and Student 5 explained that translating titles requires "a large knowledge". Student 11 mentioned that translating titles was time-consuming. Student 29 reported having difficulty translating the words that were repeated several times.

The words "I" and "he" repeat many times and I have to write it many times. (Student 29, Writ. 1)

After the workshops, many students also reported problems in translating or deciding on appropriate Vietnamese translation words for vocabulary and expressions, although they had no difficulty in comprehending their meanings.

I understand meaning of some phrases but I don't know to describe how to exact. (Student 18, Writ. 2)

Even though the number of problems related to vocabulary and expressions decreased from 82 to 70, it remained the most commonly raised problem in Translation Task 2 after the workshops. However, the students noted fewer problems in comprehending vocabulary and expressions after the workshops, with only 20 comprehension problems raised compared with 34 before the workshops. This may also indicate that they learned that translation is not all about vocabulary, or they did not consider vocabulary comprehension problems worth noting even though they may still have had such problems. The students also rarely explained their difficulties in translating vocabulary and expressions before and after the workshops.

4.1.4 Difficulties related to sentences and sentences structures

Students also reported difficulties with sentences and sentence structures. They had difficulty in comprehending and translating the sentences. They discussed problems with the length of sentences or specific types of sentence structures, such as relative clauses or imperative structures. This category also covered students' mention of grammatical points in sentences, such as tenses and comparatives. The problems related to sentences and sentence structures included those in:

- ST comprehension: The students had difficulty in understanding sentences and sentence structures.

- TT production: The students had difficulty in translating or finding Vietnamese translation words for sentences.

- Both ST comprehension and TT production: The students mentioned both kinds of problems (comprehension and production) at the same time in one comment.

The total number of responses from the students before the workshops was 13, which was slightly lower than the number reported after the workshops: 18. Firstly, there were similarly low numbers of problems cited in comprehension before and after the workshops (3 items before and 2 after). Before the workshops, Students 2 and 6 mentioned their difficulties comprehending sentence meanings without offering reasons for their lack of comprehension.

I don't understand the implication of some sentences in the text such as "There was… a bare floor". (Student 6, Writ. 1)

After the workshops, Students 9 and 25, who were the only two students to write about difficulties in understanding sentences and sentence structures, elaborated a little on their lack of sentence comprehension. They referred to sentential structures (e.g., "Not a thing" and "I was on my own", Student 9) and content (Student 25).

"Anyone… who spoke to me was in his bad books". I am not sure of what the sentence says: "because these people spoke to me, they were added in his bad books or these people were in his bad book, they weren't scare and they spoke to me". (Student 25, Writ. 2)

A similar number of difficulties (9 before and 7 after) were reported in relation to TT production and both ST comprehension and TT production. Before the workshops, the students expressed their uncertainty about the stylistic aspects of sentences, referring to the type of sentence structure (e.g., imperative sentences, Student 3 and passive sentences, Student 24), and the sentence length (e.g., short and reduced sentences, Students 14 and 15). Student 3 explained that they did not know whether the sentence was a piece of advice or an order and Student 25 thought that replicating the English use of passive sentences would be strange in translating into Vietnamese.

Translating some sentences is short. (Student 14, Writ. 1)

Passive structure-… I translate them into active but it doesn't seem logical, e.g., syringes and must be disposed of in special sharp bin. (Student 25, Writ. 1)

After the workshops, some students were more attentive to the problems caused by the length and complexity of sentences. A sentence being too long or too short could pose difficulties for the students (Students 3, 26, and 20):

Some sentences make me confuse when translating because it's too long and I find it extremely hard for me to punctuate the meaning of these sentences. (Student 26, Writ. 2)

The short sentence "Not a thing". (Student 20, Writ. 2)

Some other students wrote about their problems with special structures (Student 9) and complex sentences (Students 3, 11, 20, 19, and 30). Student 19 noted a problem with a relative clause.

Sentence structure; relative clause. (Student 19, Writ. 2)

Complex/compound sentence. (Student 30, Writ. 2)

The students wrote small numbers of ST comprehension and TT production problems before and after the workshops. Even though after the workshops, the total number of students' responses about difficulties related to sentences and sentence structures was slightly higher than before the workshops, they rarely explained why the problems existed (as evidenced by the nine unspecified difficulties).

4.2 Focus shifted to non-linguistic translation problems

Before the workshops, the majority of students were focused on linguistic problems concerning vocabulary and sentence structures. They frequently noted difficulties in comprehending these features and producing target language expressions. Focusing on linguistic problems appears to be common among foreign language learners when they translate regardless of their communicative language teaching or structural language programs as evidenced in many studies including those by Lörscher (2005) and Tirkkonen-Condit (2005). These studies show that foreign language learners tend to translate local ST items, such as words and phrases, without paying much attention to the contextualised meanings which are implied by the surrounding sentences and extratextual textual features (e.g., the author and readers). Likewise, in this study, the students' greatest concern before the workshops was whether they understood words and phrases and how to render them into the target language. Accordingly, the small proportion of non-linguistic problems (those with extratextual features and general style conventions) in the total number of identified problems before the workshops (12 out of 107) was small.

However, after the workshops, the students began to shift their focus from linguistic problems to non-linguistic ones as demonstrated by the slight decrease in identified problems with vocabulary structures and the surge in those with extratextual features and general style conventions (See Table 1). One reason for the decline in the students' identified linguistic problems could be that after the workshops the students found linguistic problems self-evident at their levels and that these problems were not worth noting. Furthermore, as the students noted fewer problems in comprehending vocabulary and expressions, it is likely that they resolved some of their comprehension problems by referring to text features as illustrated in the workshops. Meanwhile, after the workshops, more students identified pragmatic problems in dealing with "green cleaning" (Title 2a) when they were concerned about whether their translations satisfied the reader's comprehension and/or the operative function of the text.

Fernández and Zabalbeascoa (2012a, 2012b) have indicated that students who reported more non-linguistic problems (or those with general style conventions) performed better in translation than those who did not. In this study, the students diverted their attention from linguistic problems to non-linguistic problems and took into account text features presented in the workshops including extratextual and intratextual features (particularly non-linguistic extratextual ones). They began to consider problems with contextualised or pragmatic meanings determined by extratextual features while translating. This indicates that common behaviours of foreign language learners in problem identification can be changed if they are more aware of text features that influence ST comprehension and TT production

Overall, while the students did not note problems and difficulties with linguistic features as frequently as before, they paid more attention to those with extratextual features and the text type. This change in the students' awareness of translation problems would possibly lead to changes in their translation solutions will be examined in the next section.

4.3 Students' translation solutions to titles

This article only focuses on comparing students' translation solutions to the Title 1a, Text 1a ("Easy Recycling, Create your own Eden") and Title 2a, Text 2a ("Easy Green Cleaning, Create your own Eden"). Before the workshops, many students replaced "Eden" which appeared in both the titles with one of many idiomatic Vietnamese equivalents found in the dictionary including "thiên đường" or "vườn địa đàng" [heaven]. However, very few students explained their choices. Student 28 produced the TT "thiên đường" for the ST "Eden". It was not clear why Student 28 chose that as a solution and nor was it clear how they understood the purpose of the ST because the student only briefly wrote "the purpose of this text, 'create your own Eden'".

After the workshops, many students also used similar Vietnamese equivalents for Eden as they did before; however, they could have considered the reader and/or other text features even though the evidence for their justifications was limited from their written comments. Typically, students avoided too much dependence on the ST form and focused on the ST message: "Vệ sinh dễ dàng - Làm cho cuộc sống dễ dàng hơn [Clean easily - Make your life easier]" (Student 26). This certainly resulted from their attention to the reader. They wrote: "I don't know how to translate the title for Vietnamese people understood by Vietnamese thought" (Student 26, Writ. 1).

Furthermore, after the workshops, the students had varied solutions to "green cleaning". Some students (5, 8, 23, 28, 29, and 30) produced the literal Vietnamese translation or kept the ST English form in their translations. Some of them possibly thought that Vietnamese readers would understand this English phrase. Students 4 and 19 commented that they retained the English ST "green cleaning" in their Vietnamese translations due to that Vietnamese communities in Australia might understand it (despite this, Student 19 used a sense-oriented translation solution in their translation).

Rather than being constrained by word-for-word translation, some students (7, 10, 11, 15, 16, 17, 18, 19, 20, 21, and 25), which was more than one-third of the 30 students, conveyed the sense of "green cleaning", and related "green cleaning" to "the environment" in their translations. Students 11, 15, and 25 related the phrase to being "environmentally friendly" while Student 7 relied on a common Vietnamese expression used to indicate a clean environment ("xanh sạch đẹp" [green, clean and beautiful]). Student 16 related "green cleaning" to "safe cleaning". Not many of these students explained their translation solutions. Only Student 11 referred to text function and the reader in producing idiomatic translation solutions that focused on the ST message as found in the published translation.

Dọn dẹp một cách thân thiện với môi trường thật dễ dàng-Hãy tạo thiên đường của riêng bạn [Easy environmentally friendly cleaning-Create your own heaven]. (Student 11, Title 2a)

Lau chùi dễ dàng và thân thiện với môi trường-Hãy tạo nên thiên đường của chính bạn [Clean in an easy and environmentally friendly way-Create your own heaven]. (Student 15, Title 2a)

Even though the students' translation solutions to Title 2a demonstrated their avoidance of word-for-word translation after the workshops, they were limited in their expression and accuracy. For example, Student 7's translation contained some redundancy (e.g., redundant words-"rất" [very] and "để" [for the purpose of]). Other students (e.g., Students 13 and 21) distorted the meaning of the title. For example, in dealing with "green cleaning", Student 13 translated the phrase as "dọn dẹp cây cỏ" [bush and grass clearing] which did not convey the meaning of the original. Some students translated "green cleaning" literally and omitted or distorted its meaning, possibly due to their lack of ability or their own strategy.

Regardless of the students' limited expression and accuracy, more than one-third of the 30 students produced translation solutions that focused on the ST message of "green cleaning" after the workshops. The analysis of the titles showed some students' attempts to avoid word-for-word translation and produce more idiomatic translation solutions, paying attention to extratextual features including text function and reader. Some also referred to these features in their word-for-word translation or ST form retention. Generally, the students began to produce informed and functionally appropriate translations regardless of their types of translation. This change in the students' translation production was possibly due to their increased attention to the problems with extratextual features including text function and the reader.

5. Conclusion

The study attempted to incorporate elements of the functional approach to translation in a series of workshops to enhance students' awareness of translation problems. The study particularly explored differences in students' identification of translation problems before and after the workshops. The results demonstrated a moderate change in their elaboration on translation problems and difficulties. The greatest proportion of the commentary was devoted to linguistic problems with vocabulary and expressions in both translation tasks. Notably, vocabulary comprehension problems had a considerable reduction, which contributed to the slight decrease in difficulties concerning vocabulary and expressions. In fact, the amount of commentary in other categories of problems increased significantly, especially the non-linguistic problems (extralinguistic problems and difficulties related to extratextual features and general style conventions). The non-linguistic problems accounted for a small number in Translation Task 1 before the workshops. However, they increased by approximately three to four times in the post-workshop translation task, with the extratextual features-related problems becoming the second-largest category. The students' consideration of difficulties related to sentences and sentence structures did not change much. Despite this, after the workshops some students related their problems to not only sentence structure but also to content (an intratextual, non-linguistic feature).

The change in students' problem identification might be strongly linked with the fact that nearly one-third of the students produced translations based on informed decisions in consideration of text features, particularly extratextual ones. Despite the limited data on students' translation solutions, it was possible to claim that after the workshops, some students were able to deal with linguistic and extralinguistic difficulties (e.g., those with "green cleaning", Title 2a) by presenting translation solutions that incorporated what they had learnt in the workshops.

It is necessary for teachers and students in translation courses to understand that they will have an understanding that language use in translation is influenced by a range of factors such as extratextual features. Students will be then less focused on the idea of language as a linguistic code. Students' sociolinguistic awareness of language needs to be enhanced so that linguistic problems are not their only focus. As a result, they can identify a wide range of problems related to extratextual features and text type conventions. The adequate attention of these problems will allow students to avoid too literal translation and make informed decisions in translating. Generally, elements of the functional approach to translation including text analysis and translation briefs should be included in tertiary English programs to develop students' strategic competence in particular and their translation competence in general.

This study is only limited to translation problem identification and the exploration of students' written reflections on their translations. An evaluation of students' performance in translation is, therefore, still missing. Future research can focus on other issues in teaching translation including students' translation competence and the link between their problem identification and translation competence.

References

Avval, S. F. (2013). Teaching translation: A look at the way it is in Iranian universities and the way it should be. *Translation Journal, 16*(3).

Chen, Q. (2010). The application of text type in non-literary translation teaching. *Translation and Interpreting Studies, 5*(2), 208-219.

Carreres, Á., & Noriega-Sánchez, M. (2011). Translation in language teaching: Insights from professional translator training. *The Language Learning Journal, 39*(3), 281-297. doi: 10.1080/09571736.2011.567356

Colina, S., & Lafford, B. A. (2018). Translation in Spanish language teaching: The integration of a "fifth skill" in the second language curriculum. *Journal of Spanish Language Teaching*, 1-14. doi: 10.1080/23247797.2017.1407127

Cook, G. (2010). *Translation in language teaching: An argument for reassessment*. Oxford: Oxford University Press.

Fernández, F., & Zabalbeascoa, P. (2012a). Correlating trainees' translating performance with the quality of their metacognitive self-evaluation. *Perspectives: Studies in Translatology, 20*(4), 463-478. doi: 10.1080/0907676X.2011.629730

Fernández, F., & Zabalbeascoa, P. (2012b). Developing trainee translators' strategic subcompetence through metacognitive questionnaires. *Meta: Journal des Traducteurs/Translators' Journal, 57*(3), 740-762.

Gile, D. (2004). Integrated problem and decision reporting as a translator training tool. *The Journal of Specialised Translation, 2*, 2-20.

Gile, D. (2009). *Basic concepts and models for interpreter and translator training* (Rev. ed.). Amsterdam: John Benjamins Publishing.

González-Davies, M. & Scott-Tennent, C. (2005). A problem-solving and student-centred approach to the translation of cultural references. *Meta: Journal des Traducteurs/Translators' Journal, 50*(1), 160-179. doi: 10.7202/010666ar

Ho, P. V. P., & Bui, P. T. K. (2013). Senior students' reports on their translation and interpretation abilities, teaching methods, and curriculum at Ho Chi Minh Open University. *Journal of Science Ho Chi Minh City Open University, 1*(6), 68-77.

Károly, A. (2014). Translation in foreign language teaching: A case study from a functional perspective. *Linguistics and Education, 25*, 90-107. doi: 10.1016/j.linged.2013.09.009

Kokkinidou, A., & Spanou, K. (2013). The didactic use of translation in foreign language teaching: A practical example. In D. Tsagari & G. Phlōros (Eds.), *Translation in language teaching and assessment* (pp. 173-190). Newcastle upon Tyne, UK: Cambridge Scholars Publishing.

Lee, V., & Gyogi, E. (2018). Cultural-specific lexis for intercultural communication: Case studies from two different classrooms. *Journal of Language, Identity & Education*, 1-15. doi: 10.1080/15348458.2017.1418357

Leonardi, V. (2010). *The role of pedagogical translation in second language acquisition: From theory to practice*. Bern: Peter Lang.

Leonardi, V., & Salvi, R. (2016). Language pedagogy and translation studies: Towards a (re)definition of translation. In Y. Gambier & L. van Doorslaer (Eds.), *Border Crossings: Translation Studies and other disciplines* (pp. 331-348). Amsterdam: John Benjamins Publishing.

Lörscher, W. (2005). The translation process: Methods and problems of its investigation. *Meta: Journal* *des Traducteurs/Translators' Journal, 50*(2), 597-608. doi: 10.7202/011003ar

Nguyen, T. T. (2010). Nhu cầu người học và nhu cầu xã hội trong đào tạo cử nhân biên–phiên dịch tiếng Anh [Learners' needs and social needs in BA training in translation and interpreting]. *Tạp chí Khoa học ĐHSP TP HCM.* (22), 28-41.

Nord, C. (1997). *Translating as a purposeful activity: Functionalist approaches explained*. Manchester, UK: St. Jerome Publishing.

Nord, C. (2005). *Text analysis in translation: Theory, methodology, and didactic application of a model for translation-oriented text analysis*. Amsterdam: Rodopi.

Pham, H. H., & Ton, N. N. H. (2007). *Đào tạo chuyên ngành biên dịch trong thời kỳ mới: thách thức lớn cho các trường đại học* [Training of translation in the new era: great challenges for universities]. Paper presented at "Hội thảo công tác dịch thuật và đào tạo cử nhân" [Conference on Translation and Training], Hanoi, Vietnam. Retrieved from https://123doc.org//document/2592959-dao-tao-chuyen-nganh-bien-dich-trong-thoi-ky-moi-thach-thuc-lon-cho-cac-truong-dai-hoc.htm

Pham, H. H., & Tran, L. T. (2013). Developing graduate knowledge and skills for the world of work: The case of the translation curriculum in Vietnam. *Language, Society and Culture.* (36), 7-17.

Saleh, N. J., & Weda, S. (2018). Indonesian poetry translation: The problem within. *The Journal of English as an International language, 13*(2.2), 64-87.

Schäffner, C., & Wiesemann, U. (2001). *Annotated texts for translation: English-German: functionalist approaches illustrated*. Clevedon, UK: Multilingual Matters

Tirkkonen-Condit, S. (2005). The monitor model revisited: Evidence from process research. *Meta: Journal des Traducteurs/Translators' Journal, 50*(2), 405-414. doi: 10.7202/010990ar

Wongranu, P. (2017). Errors in translation made by English major students: A study on types and causes. *Kasetsart Journal of Social Sciences, 38*(2), 117-122. doi: 10.1016/j.kjss.2016.11.003

NÂNG CAO KHẢ NĂNG PHÁT HIỆN VẤN ĐỀ DỊCH THUẬT CỦA SINH VIÊN NGÀNH CỬ NHÂN ANH

Nguyễn Thị Thu Hướng

Trường Đại học Ngoại ngữ, Đại học Đà Nẵng
131 Lương Nhữ Hộc, Cẩm Lệ, Đà Nẵng, Việt Nam

Tóm tắt: Việc dạy môn dịch ở Việt Nam và một số nước thường hay tập trung vào các vấn đề ngôn ngữ và ít lưu ý đến quá trình dịch của sinh viên. Dựa trên cách tiếp cận chức năng trong dịch thuật, bài báo này nghiên cứu khả năng xác định vấn đề dịch thuật (thể hiện một phần năng lực dịch thuật) của sinh viên tham dự một khóa học dịch thuật tại một chương trình cử nhân tiếng Anh ở Việt Nam. Bài báo trình bày sự khác biệt trong khả năng xác định vấn đề dịch thuật của sinh viên thông qua việc so sánh nhận xét của sinh viên sau khi làm bài dịch và các giải pháp dịch thuật của sinh viên trước và sau các workshop về các vấn đề cơ bản trong dịch thuật. Kết quả nghiên cứu cho thấy sau workshop, sinh viên bắt đầu chú ý đến các vấn đề phi ngôn ngữ thay vì chủ yếu quan tâm các vấn đề ngôn ngữ như trước workshop. Các vấn đề phi ngôn ngữ bao gồm các vấn đề liên quan đến yếu tố ngoại văn bản và phong cách văn bản. Theo kết quả phân tích các giải pháp dịch tựa đề, sinh viên bắt đầu tránh cách dịch từ theo từ và dịch một cách phù hợp hơn. Bài báo đề xuất cần đưa các hoạt động phân tích văn bản và xác định vấn đề dịch thuật vào các khóa giảng dạy dịch thuật tại các trường ngoại ngữ.

Từ khóa: giảng dạy dịch thuật, quá trình dịch, vấn đề dịch thuật, xác định vấn đề dịch thuật, phân tích văn bản

INVESTIGATING COMPLIMENT RESPONSE STRATEGIES IN AMERICAN ENGLISH AND VIETNAMESE UNDER THE EFFECT OF SOCIAL STATUS

Nguyen Thi Thuy Linh*

VNU University of Languages and International Studies,
Pham Van Dong, Cau Giay, Hanoi, Vietnam

Received 10 March 2020
Revised 8 June 2020; Accepted 25 July 2020

Abstract: The present study seeks to investigate the effect of the social status on the use of compliment response (CR) strategies in American English and Vietnamese. To this end, two sets of data were collected with the help of a discourse completion task (DCT) illustrating twelve situational settings in which compliments were produced by ones of higher, equal, and lower status with the informants. Statistical analysis provides descriptive statistics results in terms of CR strategies on macro- and micro-level, i.e. these findings demonstrate the CR strategies of acceptance, amendment, non-acceptance, combination, and opting out. Furthermore, inferential statistics have revealed if there is a global standard in the use of CRs between American and Vietnamese native speakers. Finally, the results suggested a significant effect for the treated intervening social variable of status in determining the type of CRs.

Keywords: compliment, compliment response, social status

1. Introduction

Complementing behavior is a universal linguistic phenomenon. As a speech act which happens with a high frequency in our daily life, it plays a significant communicative function and serves to establish, consolidate, and promote interpersonal relationships (Holmes, 1988). A proper complementing behavior can make people closer and more harmonious. Being an adjacency pair, a compliment and a compliment response (CR) coexist. The responses to the compliment vary due to the social and individual elements. Different cultural customs, communicative topics, social power, gender, and educational background, etc. will affect compliment responses.

To explore compliment responses used by American and Vietnamese native speakers under the influence of social status factor, the study intends to answer the following question: *How does status affect the choices of compliment response strategies in both American and Vietnamese groups of native informants?*

2. Literature review

Compliment responding is considered the speech act that has attracted the most abundant studies in the field of pragmatics. Early work on CR research concentrated on different

* Tel: 84-362328288
Email: nthithuylinh88@gmail.com

varieties of English: American English by Herbert (1986, 1990), Manes (1983), Pomerantz (1978, 1984) and Wolfson (1983); South African English by Herbert (1989), and New Zealand English by Holmes (1988). These pioneering studies have revealed much about the various facets of both compliments and CRs: the things that are most likely to be complimented on, the kinds of interlocutors that one is likely to make compliments to, and the syntactic structures that are most often used in English for compliments and CRs, and the pragmatics of CR strategies adopted in each of these English-speaking communities.

Serious attention began to be given to CRs in other languages and cultures beginning from the 1990s. While a comprehensive review of research on compliments and CRs is seen in Chen (2010), the following sampler provides a glimpse of this vast amount of literature: Nigerian English by Mustapha (2004); Polish by Herbert (1991) and Jaworski (1995); German by Golato (2002); Spanish by Lorenzo-Dus (2001); Turkish by Ruhi (2006); Persian by Sharifian (2005); Jordanian Arabic by Farghal and Al-Khatib (2001) and Migdadi (2003); Kuwaiti Arabic by Farghal and Haggan (2006); Syrian Arabic by Nelson et al. (1996); Japanese by Daikuhara (1986), Baba (1999), Fukushima (1990), and Saito and Beecken (1997); Korean by Han (1992); Thai by Gajaseni (1995); and Chinese by Chen (1993), Yu (2004), Spencer-Oatey and Ng (2001), Yuan (2002), and Tang and Zhang (2009), among others.

These studies have discovered many subtleties and nuances about the similarities and differences among this rich diversity of languages. Speakers of German, for instance, are not found to use appreciation tokens (e.g., ‘‘Thank you’’) in CRs, although they accept compliments as much as do Americans (Golato, 2002). In Thai, social status is found to be a factor influencing speakers’ CR behavior: a compliment that flows from someone in higher social status to someone in lower social status is more likely to be accepted than one that flows in the opposite direction (Gajaseni, 1995). Instances of ‘‘impoliteness’’ are found in the Turkish data, whereby the complimenter explicitly challenges the assumption of the compliment (Ruhi, 2006, p. 70). Arabic speakers, on the other hand, are found to routinely ‘‘pay lip-service’’ (Farghal and Haggan, 2006, p. 102) to the complimenter, using a set of formulaic utterances to offer the object of the compliment to the complimenter without meaning it. In addition, gender-based differences in CRs have been attested in a number of languages. Herbert (1990), for example, finds that compliments delivered by American males are twice likely to be accepted than those delivered by females and females are twice likely to accept compliments than are males.

The diversity of findings in the literature on CRs is mirrored by the diversity of theoretical orientations these researchers adopt. Early work on CRs was informed by ethnography, sociolinguistics, sociology, and conversation analysis. Beginning from Holmes (1988), theories of politeness began to be used by researchers to account for their findings. These politeness theories, particularly Brown and Levinson’s theory, have been the dominating theoretical framework for CR researchers, although not all of them have been found adequate (e.g., Chen, 1993; Ruhi, 2006).

Recent years have seen proposals of new theoretical constructs in CR research. Sharifian (2005) explains Persian CRs in terms of cultural schemas, arguing that Persian CRs are motivated by the schema of shekasteh-nafsi ‘‘broken self,’’ glossed as ‘‘modesty’’ or "humility." Finding classical theories wanting in their explanatory adequacy to inform CR's in Turkish, Ruhi (2006) proposes the notion of self-politeness-based on but different from Chen's (2001) model of self-politeness—which includes three aspects: display confidence, display individuality, and display impoliteness. Ruhi and Doğan (2001), on the other hand, posit that Sperber and Wilson (1993) theory of relevance is a viable alternative to account for the cognitive processing of compliments and CRs in Turkish.

Researchers in CR research have also adopted a range of taxonomies for categorizing CR utterances. Pomerantz's (1978, p. 81–82) seminal work on CR identifies two conflicting constraints facing a compliment responder:

A. Agree with the complimenter

B. Avoid self-praise

Constraint A explains acceptance of compliments, often expressed by appreciation tokens (e.g., "Thank you"). Constraint B motivates those strategies that downgrade the value of the objects of compliments (e.g., "That's a beautiful sweater!" "It keeps out the cold") or to shift the credit away from the responder herself (e.g., "That's a beautiful sweater!" "My best friend gave it to me on my birthday"). These two general principles are refined into three categories in Herbert (1986): Agreement, Non-Agreement, and Other Interpretations. Under each of these three categories are several subtypes of responses. While this taxonomy has been popular, it has not been the only one. Holmes' (1988) system of classification, for example, is clearly different, whereby she classifies 12 types of CRs - labeled differently from Herbert's-into three broad categories: Acceptance, Deflection/Evasion, and Rejection. Yu (2004) groups her Taiwanese CRs into six types. Yuan (2002) uses yet another system of labels for the 12 semantic formulas she has identified from her Kunming Chinese data, including two that have not been identified in previous studies: invitation and suggestion.

In spite of this wide variety of taxonomies, however, one can discern a convergence in the way CRs are categorized, that the tripartite system - Acceptance, Deflection/Evasion, and Rejection - originally proposed by Holmes (1988) and supported by Han (1992) and Chen (1993)—has been gaining currency (Ruhi, 2006; Tang and Zhang, 2009; among others). This taxonomy, first, reflects the insights of Pomerantz's (1978) constraints as seen above. The need to agree with the complimenter motivates the acceptance of a compliment; the need to avoid self-praise motivates the rejection of a compliment, while the need to strike a balance between the two constraints leads to utterances that mitigate—either deflect or evade the compliment.

To reflect the nature of the data collected, both regarding the American and Vietnamese data sets, I decided to embed some of the compliment response strategies nominated by Yu (2003). The annexation of Ruhi's taxonomy (2006) is reflected through the inclusion of the sub-category of Appreciation (token + comment,) as an acceptance strategy and addition of three combination strategies on macro-level. This macro-level strategy - Combination - accounts for the responses manifesting two sub-categories of the macro-level strategies of Acceptance, Deflection/Evasion or Rejection. The following table depicts the chosen taxonomy of compliment responses that I have adapted and employed for the analysis.

Table 1: Adapted taxonomy of Compliment responses

Macro-level strategies	*Micro-level strategies*	*Example*
I. Acceptance	*Appreciation token*	*- Thank you! (Cám ơn!)*
	Agreement	*- Yeah, it is. (Đúng vậy!)*
	Expressing gladness	*- I am so glad that I can help! (Mình rất vui vì có thể giúp được cậu!)*
	Upgrade	*- Maybe it's because I'm very active. - Damn it, I'm perfect. (Chuyện! Tao chỉ có là hoàn hảo!)*
	Joke	*- What a cute chubby little boy! - Cute as his mom and chubby as his dad! (- Ôi em bé dễ thương mũm mĩm yêu quá! - Uh, dễ thương giống mẹ còn mũm mĩm giống bố!)*
	Laughter	*You look smarter with this new laptop! – [Loud laughter] (- Có con máy mới nhìn ngon hẳn! - Haha)*
	Acceptance association	*- Thank you! I am so glad you like it! (Cám ơn! Mình rất vui vì bạn thích!)*
II. Amendment	*Return*	*- Your mother used to cook very well, too. (Mẹ bạn nấu ăn cũng rất ngon đấy!)*
	Downgrade	*- It's my duty, I do it with pleasure. (Đây là trách nhiệm của mình mà!)*
	Question	*- You look smart with the new laptop! - What do you mean to "look smart"? (Bạn trông thật bảnh với chiếc máy tính mới! - Ý bạn "bảnh" là thế nào?)*
	Comment	*- Your dress looks nice. - I bought it yesterday. (Váy đẹp nhỉ!- Mình mới mua hôm qua!)*
	Transfer	*- I couldn't have done it without you. (Nếu như không có cô, em không thể có được ngày hôm nay!)*
	Amendment association	*- Really? You think so? Honestly I just thought I was lucky. (Thật sao? Bạn nghĩ vậy ư? Thực tình mà nói mình chỉ ăn may thôi!)*

Non-acceptance	*Disagreement*	*- I don't think so. (Mình không nghĩ vậy!)*
	Qualification	*-You must be very smart. You did well on the previous exam. - Not really, you did better. (Cậu giỏi thật đấy! Bài kiểm tra hôm nọ làm siêu thật!- Không hẳn, cậu làm tốt hơn.)*
	Diverge	*- You did well on the previous exam! - Let's try to study harder and get the scholarship! (Bài kiểm tra hôm nọ cậu làm giỏi thật!- Chúng mình cùng cố gắng học hành chăm chỉ hơn để lấy học bổng nhé!)*
	Non-acceptance association	*- No, you did a better job. Why don't we get a drink after school? (Không, cậu làm tốt hơn. Chúng mình sau giờ học đi uống nước đi!)*
IV. Combination	*Combination 1 (accept+amend)*	*- Thank you. I couldn't have done it without you. (Cám ơn thầy. Em không thể được như vậy nếu không có thầy chỉ bảo.)*
	Combination 2 (accept and non-accept)	*- Pleasure was all mine. Let's study harder next term. (Đây là niềm vinh hạnh của tớ. Kì tới học hành chăm chỉ hơn nhé!)*
	Combination 3 (amend and non-accept)	*- I tried really hard to get the scholarship but honestly you deserved it more than me. (Tớ đã cố gắng rất vất vả để giành học bổng đấy nhưng kì thưc, tớ thấy cậu xứng đán hơn tớ.)*
V. Opting out	*Opting out with fillers*	*- You look great!- Awwwww (Uầy! Trông ngon đấy!)*
	Opting out without anything/ no acknowledgement (silence)	*- You look smart with the new laptop! - [Silence] (Có máy tính mới nhìn sáng sủa hẳn!- [Im lặng])*
	Opting out with topic change	*- What a nice car! – What do you think of the color? (Xe mới đẹp nhỉ!- Cậu nghĩ sao về màu sơn xe?)*
	Expressing embarrassment	*- You are so good at it! – Oops, I am embarrassed. (Giỏi quá cơ! – Ôi, ngại quá!)*

3. Methodology

3.1. Participants

The overall population of participants in this study was 237, which was divided quite evenly into two big groups- American natives and Vietnamese natives. In the American group, the number of female respondents was 61 while 56 of them were male. The Vietnamese group also had a tendency that more female informants took part in the study than male ones. Out of 120 Vietnamese participants, 68 ones were female while the number of male ones was 52.

Recruiting informants was based on two criteria that decided upon whether an informant was eligible for the research or not. Each informant was asked two questions and a positive answer to both of them qualified them as potential participants. The two criteria are those related to the country of birth and their mother tongue.

Criteria questions for recruiting informants for the study:

- Are you native speaker of American/ Vietnamese?
- Were you born in the U.S/Vietnam?

Some tendencies of how American and Vietnamese informants have been found are discovered and my considerations on this very process might be of some help to future researchers with similar research methodology criteria who will embark on the quest for study participants.

Table 2: Participants' characteristics

Speaker group	*American*	*Vietnamese*
Number of females	*61*	*68*
Number of males	*56*	*52*

3.2. Research instruments

A pilot DCT was designed and tested. The purpose of this trial run was to identify the existing flaws in the wordings and order of the questions as well as potential practical problems in following the research procedure. In particular, it tested the social variables set out in the research questions (gender social status and topics of compliments). The initial version of the DCT was distributed to a female Vietnamese PhD candidate who is an experienced TESOL practitioner as well as an English-Vietnamese proficient translator and a male American researcher in COE College who is living in Iowa. They were asked to comment on the appropriateness of the content and wording after they had finished filling it in. A Vietnamese version of this DCT was also sent to 23 second-year students of International Standard Program in Faculty of English, the University of Languages and International Studies, Vietnam National University. The responses gathered from the pilot test were used as reference for improving the final version of the DCT.

Because the DCT was first constructed in English and was later translated into Vietnamese, cultural transposition had to be considered (Blum-Kulka, House & Kasper, 1989, p. 274). Accordingly, the Vietnamese social context had to be taken into account in the process of translation. Several factors may affect the quality of the translation: the translator's linguistic competence, her knowledge of the culture and the people under study, the autobiography of those involved in the translation, and the circumstances in which the translation takes place (Temple, 1997, p. 610). The DCT, first constructed in English, was therefore translated into Vietnamese by the researcher, then a proficient bilingual translated the Vietnamese back into English for comparison with the original English version for mismatches and any changes needed to ensure conceptual equivalence.

The DCT used in this research consisted of two parts, the first one is the introduction to the survey and the second section contains 12 situations which were discreetly constructed to investigate the gender, social status and complimenting topic variables. Full versions in both languages of the DCT can be found in the Appendix.12 situations are named as in the following table:

Table 3: List of situations in the DCT Questionnaire

Situation 1:	*Thesis defense*	*Situation 7:*	*Weight loss*
Situation 2:	*Help at meal*	*Situation 8:*	*New car*
Situation 3:	*Nice outfit*	*Situation 9:*	*Scholarship*
Situation 4:	*First baby*	*Situation 10:*	*Helping friend*
Situation 5:	*Inspiring lesson*	*Situation 11:*	*New haircut*
Situation 6:	*Humorous boss*	*Situation 12:*	*New MacBook*

With an aim to investigate the social status variable, compliments in situations 1-4 are issued by complimenters of high social status to recipients of low social status. Thus, the compliment response will flow from Low (L) status to High (H) status. Compliments in situations 5-8 are issued by complimenters of low social status to recipients of high social status. That is, the compliment response will flow from High (H) status to Low (L) status. The characters chosen to represent a person of high social status included a boss at work, a supervisor, and mother-in-law. Low status characters were represented by a university student, a subordinate, a daughter/son-in-law and a nephew/niece. Compliments and compliments responses in situations 9-12 are interchanged between friends. Thus, the compliment response flows horizontally between colleagues and peers, that is, between two persons of equal social status.

Table 4: Social status distribution in the DCT questionnaire

High to low	Low to high	Equal
Situation 1	Situation 5	Situation 9
Situation 2	Situation 6	Situation 10
Situation 3	Situation 7	Situation 11
Situation 4	Situation 8	Situation 12

3.3. Data collection procedure

The DCT questionnaire was administered in person to both groups of respondents who were given adequate time to complete the surveys at their own pace. The reason behind was the fact that due to the relatively high number of open-ended questions (12 items) seeking spontaneity in providing responses would possibly touch the borders of affective factors such as stress leading to unreliable records.

Importantly, during the coding of the compliment responses, a sample of each corpus was examined by two other raters (one male and one female) to achieve inter-rater reliability. For each part, 20% of the data were randomly exposed to recoding by a second and third rater as suggested by Cohen (1960, as cited in Yu, 2005, p. 98). In this way, another sex-based confound would be remedied for through coming up with an average reliability rate of these two opposite sex-coders.

3.4. Data analysis

The DCT data will be statistically analyzed using IBM statistical software package SPSS. Data were coded for social status (higher, lower, and equal status). Social status was defined as institutionalized role (teacher, student), family role (mother, daughter-in-law, etc.), or age (senior, junior colleague. Using these distinctions as a base, I coded status as a binary-value, that is, either the addressee was higher status (+ status) or low status (- status) (cf. Yu, 2004). The coded data of the DCTs were analyzed using the Statistical Package for Social Sciences (SPSS 20) software. The Chi-square statistical tool was employed for analysis frequency distributions, chi-square goodness-of-fit test, cross-tabulations, and tests of significance. It also allowed investigation of the possible influence of social status on the CR choices of strategies, as well as the 2- cohort-types of respondents and their choice of macro-level strategies and micro-level strategies used to respond to compliments. The standard of P <0.05 was used to show the significance level.

4. Results and discussion

4.1. Analysis of American CRs

The status variable has been controlled in the very process of designing DCT questionnaire and data collection in order to obtain data that are balanced and all variable values i.e. high, low, equal are proportionally assigned to data points. Therefore, the chi-square goodness-of-fit test was skipped since the numbers have been equally distributed (468 for each). This indicates that further analysis taking the status variable into consideration will provide unbiased results based on a representative sample. Status-based results are also exhibited on two levels- macro and micro-level.

Table 5 highlights adjusted residuals which explain that compliments given by someone of higher status are tended to be more accepted (adjusted residual= 2.6) and ones given by someone of equal status are less accepted (adjust residual= -3.0). Whereas, non-acceptance is found to be overrepresented between status equals i.e. friends or classmates (adjusted residual= 4.9) and underrepresented with someone from higher status (adjusted residual= -3.9).

Table 5: Contingency table for macro-level strategies and status relation (American data)

*Macrolevel strategies * Status Crosstabulation*						
			Status			*Total*
			Lower	*Higher*	*Equal*	
Macrolevel strategies	*Acceptance*	*Count*	*256*	*275*	*226*	*757*
		Adjusted Residual	*.4*	*2.6*	*-3.0*	
	Amendment	*Count*	*74*	*89*	*80*	*243*
		Adjusted Residual	*-1.0*	*1.2*	*-.1*	
	Non-acceptance	*Count*	*53*	*36*	*88*	*177*
		Adjusted Residual	*-1.0*	*-3.9*	*4.9*	
	Combination	*Count*	*78*	*62*	*64*	*204*
		Adjusted Residual	*1.6*	*-1.0*	*-.6*	
	Opting out	*Count*	*7*	*6*	*10*	*23*
		Adjusted Residual	*-.3*	*-.7*	*1.0*	
Total		*Count*	*468*	*468*	*468*	*1404*

Chi-Square Tests			
	Value	*df*	*Asymp. Sig. (2-sided)*
Pearson Chi-Square	*33.441*[a]	*8*	*.000*
Likelihood Ratio	*32.996*	*8*	*.000*
Linear-by-Linear Association	*1.627*	*1*	*.202*
N of Valid Cases	*1404*		

a. 0 cells (0.0%) have expected count less than 5. The minimum expected count is 7.67.

The analysis of compliment responses in terms of the status relation between a complimenter and a complimentee reveals some differences among three groups of categorization. Firstly, acceptance is still the most favored strategy out of 5 macro-level ones; however, when commuting with status equals, respondents chose only 48.30% whereas respondents of higher status chose 58.60%. Another remarkable difference is situated on the choice of non-acceptance. This macro-level strategy was favored more when it comes to communicating with people of equal status.

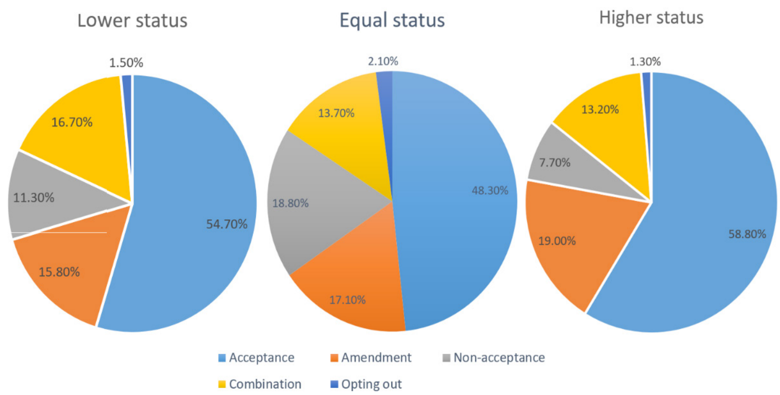


Figure 1: Compliment responses on macro-level across status relations (American data)

When it comes to micro-level strategies, interesting results are found in the choice of some strategies namely expres2sing gladness, acceptance association, return, comment, disagreement, combination 1, combination 3, and opting out with fillers. As can be seen from the highlights in table 6, there is a small overrepresentation of acceptance association (in people of lower status), return (in people of lower status), comment (in people of higher status), combination 1 (in people of lower status) and opting out with fillers (in people of equal status). This is reflected by the number of adjusted residuals ranging from 2.0-3.0. In the choice of expressing gladness, people of higher status tended to make much more use of this strategy (adjusted residual= 6.1) while people of equal status used much less than expected (adjusted residual=-5.6). What's more, the great gap is also witnessed in the choice of disagreement strategy. When communicating with status equals (adjusted residual= 5.5), respondents chose to reply no more than when communicating with one of higher status (adjusted residual= -5.0).

Table 6: Contingency table of micro-level strategies and informants' status (American data)

*Micro-level strategies * Status Crosstabulation*						
			Status			*Total*
			Lower	*Higher*	*Equal*	
Micro-level strategies	*Appreciation token*	*Count*	*136*	*143*	*139*	*418*
	Agreement	*Adjusted Residual*	*-.4*	*.5*	*.0*	
		Count	*24*	*18*	*31*	*73*
	Expressing gladness	*Adjusted Residual*	*-.1*	*-1.6*	*1.7*	
		Count	*36*	*68*	*11*	*115*

	Adjusted Residual	*-.5*	*6.1*	*-5.6*	
Upgrade	*Count*	*16*	*17*	*10*	*43*
Joke	*Adjusted Residual*	*.5*	*.9*	*-1.4*	
	Count	*16*	*14*	*21*	*51*
Laughter	*Adjusted Residual*	*-.3*	*-.9*	*1.2*	
	Count	*1*	*0*	*0*	*1*
Acceptance association	*Adjusted Residual*	*1.4*	*-.7*	*-.7*	
	Count	*27*	*14*	*14*	*55*
Return	*Adjusted Residual*	*2.5*	*-1.3*	*-1.3*	
	Count	*20*	*10*	*11*	*41*
	Adjusted Residual	*2.1*	*-1.2*	*-.9*	
Downgrade	*Count*	*5*	*12*	*9*	*26*
	Adjusted Residual	*-1.5*	*1.4*	*.1*	
Question	*Count*	*17*	*14*	*18*	*49*
	Adjusted Residual	*.2*	*-.7*	*.5*	
Comment	*Count*	*24*	*49*	*38*	*111*
	Adjusted Residual	*-2.7*	*2.5*	*.2*	
Transfer	*Count*	*8*	*4*	*2*	*14*
	Adjusted Residual	*1.9*	*-.4*	*-1.5*	
Amendment association	*Count*	*0*	*0*	*2*	*2*
	Adjusted Residual	*-1.0*	*-1.0*	*2.0*	
Disagreement	*Count*	*43*	*19*	*74*	*136*
	Adjusted Residual	*-.4*	*-5.0*	*5.5*	
Qualification	*Count*	*5*	*7*	*6*	*18*
	Adjusted Residual	*-.5*	*.5*	*.0*	
Diverge	*Count*	*5*	*9*	*6*	*20*
	Adjusted Residual	*-.8*	*1.1*	*-.3*	
Non-acceptance association	*Count*	*0*	*1*	*2*	*3*
	Adjusted Residual	*-1.2*	*.0*	*1.2*	
Combination 1	*Count*	*63*	*48*	*38*	*149*
	Adjusted Residual	*2.5*	*-.3*	*-2.1*	
Combination 2	*Count*	*9*	*13*	*12*	*34*
	Adjusted Residual	*-.9*	*.6*	*.2*	
Combination 3	*Count*	*6*	*1*	*14*	*21*
	Adjusted Residual	*-.5*	*-2.8*	*3.3*	
Opting out with fillers	*Count*	*0*	*1*	*5*	*6*
	Adjusted Residual	*-1.7*	*-.9*	*2.6*	
Opting out with silence	*Count*	*4*	*2*	*2*	*8*
	Adjusted Residual	*1.0*	*-.5*	*-.5*	
Opting out with topic change	*Count*	*1*	*0*	*1*	*2*
	Adjusted Residual	*.5*	*-1.0*	*.5*	
Expressing embarrassment	*Count*	*2*	*4*	*2*	*8*
	Adjusted Residual	*-.5*	*1.0*	*-.5*	
Total	*Count*	*468*	*468*	*468*	*1404*

4.2. Analysis of Vietnamese CRs

Like American data, Vietnamese data are balanced when the status variable is tested, for the DCT questionnaire and data collection instrument were made with a view to obtaining the data that would show somewhat equally distributed status values (lower, higher, equal) i.e. status variable was a controlled one and hence, the data yielded proportional numbers of data points for each variable values. As such, the step to conduct the chi-square goodness-of-fit test was skipped since the test statistics containing a high p-value is rested assured.

The results on the influence of status onto the choice of a compliment response strategy made by Vietnamese informants will be presented from the perspectives of macro- and micro-level strategies.

A look at the chi-square test table in table 7 reveals that status does exert an influence on the choice of macro-level strategies among Vietnamese native speakers. The p-value is 0.00, which is smaller than the significance value of 0.05. As the adjusted residuals imply, there some important differences in responding to compliments when status of the interlocutors engaged in a complimenting event is analyzed. It was discovered that people of lower status tended to accept compliments more frequently than it was expected (adjusted residual= 3.8). Meanwhile, one of higher status did not choose to do so (adjusted residual= -2.2). In terms of amendment and combination strategies, higher status respondents showed a higher than expected preference towards those (adjusted residual for amendment=4.2, adjusted residual for combination=2.3). In contrast, they showed a reluctance to choose non-acceptance and opting out strategies since the adjusted residuals for both are below -2.0. Regarding the group of status equals, it is observable that they did not very often go for amendment (adjusted residual=-4.4) and combination (adjusted residual=-2.1). In the meantime, there was a great overrepresentation of non-acceptance (adjusted residual=7.4) and opting out strategy (adjusted residual=2.4) among status equals. This may be reasoned by the fact that when communicating with friends or classmates of equal status, respondents are more at ease to turn in their refusal without fearing to lose face of others.

Table 7: Contingency table for macro-level strategies and status relation (Vietnamese data)

*Macro-level strategies * Status Crosstabulation*						
			Status			*Total*
			High	*Low*	*Equal*	
Macro-level strategies	*Acceptance*	*Count*	*136*	*186*	*140*	*462*
		Expected Count	*154.0*	*154.0*	*154.0*	*462.0*
		Adjusted Residual	*-2.2*	*3.8*	*-1.7*	
	Amendment	*Count*	*158*	*126*	*90*	*374*
		Expected Count	*124.7*	*124.7*	*124.7*	*374.0*
		Adjusted Residual	*4.2*	*.2*	*-4.4*	
	Non-acceptance	*Count*	*76*	*64*	*149*	*289*
		Expected Count	*96.3*	*96.3*	*96.3*	*289.0*
		Adjusted Residual	*-2.8*	*-4.5*	*7.4*	
	Combination	*Count*	*95*	*78*	*66*	*239*
		Expected Count	*79.7*	*79.7*	*79.7*	*239.0*
		Adjusted Residual	*2.3*	*-.3*	*-2.1*	
	Opting out	*Count*	*15*	*26*	*35*	*76*
		Expected Count	*25.3*	*25.3*	*25.3*	*76.0*
		Adjusted Residual	*-2.6*	*.2*	*2.4*	
Total		*Count*	*480*	*480*	*480*	*1440*
		Expected Count	*480.0*	*480.0*	*480.0*	*1440.0*

Chi-Square Tests			
	Value	*df*	*Asymp. Sig. (2-sided)*
Pearson Chi-Square	*85.782[a]*	*8*	*.000*
Likelihood Ratio	*83.790*	*8*	*.000*
Linear-by-Linear Association	*3.444*	*1*	*.063*
N of Valid Cases	*1440*		
a. 0 cells (0.0%) have expected count less than 5. The minimum expected count is 25.33.			

Figure 2 brings in a brief summary of Vietnamese choices on macro-level strategies with a view to compare and contrast among three groups of status categorization. As clearly seen from the pie charts, the ranks of 5 macro-level strategies are quite differentiated. In terms of the differences, two points of contrast can be recognized. Firstly, the choice of amendment strategy within the group of higher status, which accounts for the biggest share of all, is surprisingly higher than that of the status equals (32.92% and 18.75% respectively). Another difference lies on the choice of non-acceptance strategy. In comparison with the other groups, status equals were more inclined to decline a compliment. Their percentage of non-acceptance strategy almost doubles those of lower status and higher status groups. (31.04% vs. 13.33% and 15.83%).

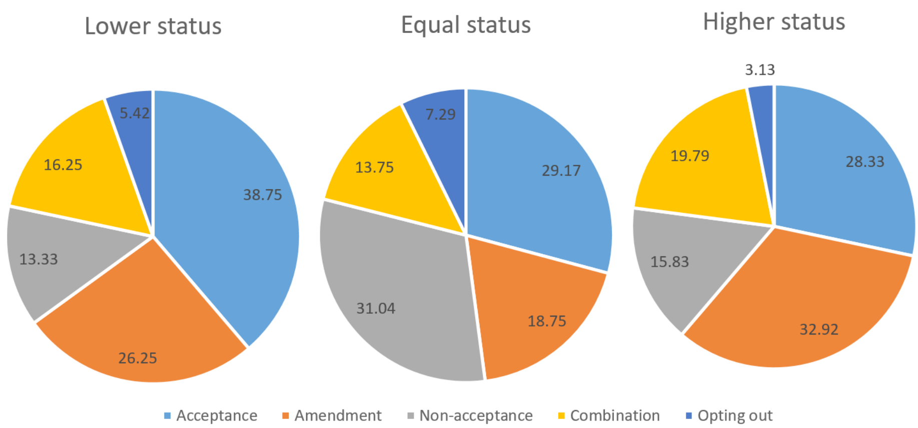


Figure 2: Compliment responses on macro-level across status relation (Vietnamese data)

As regards micro-level strategies and their distribution with respect to status factor, some intriguing and thought-provoking results were obtained (see table 8). First of all, the p-value is smaller than the significance level of 0.05, which reveals a great statistical influence of status relation on the choices of micro-level strategies. This means that the great differences can be found within 24 sub strategies.

When responding to compliments, ones of lower status were more inclined to shift credit to the complimenters or to the third party (adjusted residual=10.8) whereas they made less use of diverge and question strategy (adjusted residual= -4.0 and -4.4, respectively). This may be seen as an act of politeness from the Vietnamese complimentees since just saying thanks and asking someone of higher status questions may be regarded as being rude.

Regarding the response choices of Vietnamese respondents who were at a higher social distance, it is proven that they rarely chose to reject the compliments (adjusted residual= -7.5) or shift credit to other (adjust residual=-5.0). On the contrary, they were more at ease to opt for sub strategies such as giving comment (adjusted residual=3.8), diverging the compliments (adjusted residual= 3.8) or making jokes (adjusted residual=2.9).

Lastly, in stark contrast to the choices from ones from lower and higher status, only Vietnamese status equals showed a consuming preference towards the disagreement strategy (adjusted residual= 8.5) and a certain predilection for asking questions (adjusted residual=4.0). Perhaps, since the interlocutor was friend or classmate, the respondents felt more comfortable to express their inner feelings by rejecting the compliments without the fear of being judged or embarrassing the complimenters.

Table 8: Contingency table of micro-level strategies and informants' status (Vietnamese data)

*Micro-level strategies * Status Crosstabulation*						
			Status			*Total*
			Lower status	*Higher status*	*Equal*	
Micro-level strategies	*Appreciation token*	*Count*	*73*	*83*	*62*	*218*
		Adjusted Residual	*.1*	*1.6*	*-1.7*	
	Agreement	*Count*	*11*	*23*	*22*	*56*
		Adjusted Residual	*-2.2*	*1.3*	*1.0*	
	Expressing gladness	*Count*	*6*	*16*	*3*	*25*
		Adjusted Residual	*-1.0*	*3.3*	*-2.3*	
	Upgrade	*Count*	*21*	*16*	*14*	*51*
		Adjusted Residual	*1.2*	*-.3*	*-.9*	
	Joke	*Count*	*3*	*22*	*15*	*40*
		Adjusted Residual	*-3.5*	*2.9*	*.6*	
	Laughter	*Count*	*4*	*5*	*5*	*14*
		Adjusted Residual	*-.4*	*.2*	*.2*	
	Acceptance association	*Count*	*18*	*21*	*19*	*58*
		Adjusted Residual	*-.4*	*.5*	*-.1*	
	Return	*Count*	*15*	*14*	*15*	*44*
		Adjusted Residual	*.1*	*-.2*	*.1*	
	Downgrade	*Count*	*24*	*26*	*26*	*76*
		Adjusted Residual	*-.3*	*.2*	*.2*	
	Question	*Count*	*2*	*17*	*29*	*48*
		Adjusted Residual	*-4.4*	*.3*	*4.0*	
	Comment	*Count*	*49*	*63*	*18*	*130*
		Adjusted Residual	*1.1*	*3.8*	*-4.9*	
	Transfer	*Count*	*61*	*3*	*0*	*64*
		Adjusted Residual	*10.8*	*-5.0*	*-5.8*	
	Amendment association	*Count*	*7*	*3*	*2*	*12*
		Adjusted Residual	*1.8*	*-.6*	*-1.2*	
	Disagreement	*Count*	*61*	*20*	*119*	*200*
		Adjusted Residual	*-.9*	*-7.5*	*8.5*	

	Qualification	*Count*	*4*	*4*	*1*	*9*
		Adjusted Residual	*.7*	*.7*	*-1.4*	
	Diverge	*Count*	*9*	*40*	*26*	*75*
		Adjusted Residual	*-4.0*	*3.8*	*.3*	
	Non-acceptance association	*Count*	*2*	*0*	*3*	*5*
		Adjusted Residual	*.3*	*-1.6*	*1.3*	
	Combination 1	*Count*	*63*	*39*	*23*	*125*
		Adjusted Residual	*4.2*	*-.5*	*-3.7*	
	Combination 2	*Count*	*7*	*27*	*19*	*53*
		Adjusted Residual	*-3.2*	*2.8*	*.4*	
	Combination 3	*Count*	*25*	*12*	*24*	*61*
		Adjusted Residual	*1.3*	*-2.3*	*1.0*	
	Opting out with fillers	*Count*	*0*	*0*	*1*	*1*
		Adjusted Residual	*-.7*	*-.7*	*1.4*	
	Opting out with silence	*Count*	*4*	*10*	*16*	*30*
		Adjusted Residual	*-2.3*	*.0*	*2.3*	
	Opting out with topic change	*Count*	*4*	*7*	*10*	*21*
		Adjusted Residual	*-1.4*	*.0*	*1.4*	
	Expressing embarrassment	*Count*	*7*	*9*	*8*	*24*
		Adjusted Residual	*-.4*	*.4*	*.0*	
Total		*Count*	*480*	*480*	*480*	*1440*

4.3. Discussion

The significant values (smaller than 0.01) from the chi-square test in both American and Vietnamese native groups of informants indicate that social status exerts a certain impact on the choices of CR strategies. However, the influence of social status on each group's CR choices is quite varied.

In terms of the American group, compliment receivers of higher status tended to accept more than deny a positive comment. In contrast, it is more frequent among equal interactions to turn down a compliment than to accept. This seems to correlate with the Politeness theory proposed by Brown and Levinson (1987). To be specific, people choose to accept the compliment given by ones of lower status in order to save face for the compliment giver and in other words, to preserve the harmony and maintain the positive face during communicative exchanges. In terms of social equals, since there exists a certain degree of intimacy with each other, they found it easier to express themselves and maintain their negative face without the fear of being misunderstood as being impolite or rude.

Regarding the impact of social status on the Vietnamese group, it is observable that the Vietnamese reacted quite differently when the communicating partners were ones of different status. Notably, ones of higher status made use of amendment and combination more often than acceptance and non-acceptance. On the contrary, ones of lower status more frequently chose to accept the compliment. Finally, towards social equals, Vietnamese informants were found to unexpectedly turn down a compliment rather than amend it. To reason for that, it is important to look back at the influence of Chinese culture on ours. For such a long time, Vietnamese culture is deeply rooted by the Confucian ideology from China, especially the Five Constant Virtues including Kindness, Decorum, Uprightness, Wisdom and Faithfulness, among which Uprightness implies the respectfulness towards ones of higher status and harmony with ones of equal and lower status during social interactions. Thus, it may seem awkward or somehow rude if they just accept or turn down the positive comments from ones of higher status. In this regard, this reasons for the fact that acceptance is by far not the most preferred strategy of Vietnamese respondents. However, in case of equal status interaction, it seems that the Vietnamese informants are less constrained to react to compliments. One evidence is that the non-acceptance occurred at the highest rate out of the five macro-level strategies. Using rejecting strategies is considered an instance of adhering to the Modesty Maxim as based on Leech (1983), which means that the complimentee tries to attend to his or her own positive face so that his or her behavior can be regarded as polite and not to save the complimenter's positive face.

Concerning the relative social factors of distance and power, the politeness systems suggested by Scollon and Scollon (2001) could apply to all the recorded response types.

The assumptions of unequal power status and distant relations of the hierarchical politeness system set the ground for the respondents to suggest acceptance, non-acceptance, and amendment strategies. In such almost alien contexts the respondents behaved differently in their application of compliment response categories. First, the lower status addresses tried to establish "common grounds" (Chen, 1993, p. 58) with their speakers giving their gladness accounts or express their gratitude via thanking them. Commenting was also used in addressing the superordinates to slightly decrease the worth of the object by an emphasis on the role of factors like a long period of exercising, the price, etc. Then, to agree with the complimenter and avoid self-praise at the same time the subordinate respondents took refuge in downgrade and qualification response categories to indicate two things: either to play down the value of the complimented objects by referring to their defects or to suggest the praised trait as merely a requirement of their tasks not a sort of natural talent.

The deferential politeness system encompassed equal social distance in the interactional contexts. In such situations the participants are supposed to suggest their responses out of respect for their addresser. With this background, the complimentees frequently tended to return the complimentary force to the speaker in an attempt to maximize benefit to him/her, too, in line with the earlier noted Tact Maxim. Furthermore, such respect in a strong sense could even lead the recipients to disagree with the speaker in spite of the truth of the compliment.

The third solidarity politeness system entailed equal, close relations between the interlocutors out of which neutral elaboration major response category was raised. Questioning the truth of the statements as well as assigning the reasons of the success to other third person forces resulted from this category in the respondents' attempts to avoid self-praise.

5. Conclusion

The present study tried to contribute to the existing literature on speech acts, in general and compliment responding, in particular. A focus on the sociolinguistic variable of status in performing such illocutionary acts proved helpful in examining the unnoticed or less attended-to-corners of the earlier studies. As to the social distance concerns, acceptance response category was grounded in hierarchical politeness system while return rooted in the deferential politeness system and solidarity politeness system raised the non-acceptance response category.

Although through this study, the attempt was made to select the participants from different parts of both countries, the results cannot be generalized to all the American and Vietnamese native speakers with certainty. Meanwhile, the unveiled complexities might help the speakers from both cultures in selection of their compliment responses.

References

Baba, J. (1999). *Interlanguage pragmatics: Compliment responses by learners of Japanese and English as a second language*, (Vol. 4). Lincom Europa.

Blum-Kulka, S., House, J., & Kasper, G. (1989). *Cross-cultural pragmatics: Requests and apologies* (Vol. 31). Ablex Pub.

Brown, P., Levinson, S. C., & Levinson, S. C. (1987). *Politeness: Some universals in language usage (Vol. 4)*. Cambridge university press.

Chen, R. (1993). Responding to compliments A contrastive study of politeness strategies between American English and Chinese speakers. *Journal of Pragmatics*, *20*(1), 49-75.

Chen, R. (2001). Self-politeness: A proposal. *Journal of pragmatics*, *33*(1), 87-106.

Chen, R. (2010). A cross-cultural survey of research on complimenting and compliment responding. *Handbook of Pragmatics, 7*.

Daikuhara, M. (1986). A study of compliments from a cross-cultural perspective: Japanese vs. American English. *Working Papers in Educational Linguistics* (*WPEL*), *2*(2), 6.

Farghal, M., & Al-Khatib, M. A. (2001). Jordanian college students' responses to compliments: A pilot study. *Journal of Pragmatics*, *33*(9), 1485-1502.

Farghal, M., & Haggan, M. (2006). Compliment behaviour in bilingual Kuwaiti college students. *International Journal of Bilingual Education and Bilingualism*, *9*(1), 94-118.

Fukushima, N. J. (1990). A study of Japanese communication: compliment-rejection production and second language instruction (Doctoral dissertation, University of Southern California).

Gajaseni, C. (1995). *A contrastive study of compliment responses in American English and Thai including the effect of gender and social status* (Unpublished doctoral dissertation). University of Illinois.

Golato, A. (2002). German compliment responses. *Journal of Pragmatics*, *34*(5), 547-571.

Han, C. H. (1992). A Comparative Study of Compliment Responses: Korean Females in Korean Interactions and in English Interactions. Working Papers in Educational Linguistics, *8*(2), 17-31.

Herbert, R. K. (1986). Say" thank you"-or something. *American speech*, *61*(1), 76-88.

Herbert, R. K. (1990). Sex-based differences in compliment behavior 1. *Language in Society*, *19*(2), 201-224.

Herbert, R. K. (1991). The sociology of compliment work: An ethnocontrastive study of Polish and English compliments. *Multilingua-Journal of Cross-Cultural and Interlanguage Communication*, *10*(4), 381-402.

Herbert, R. K., & Straight, H. S. (1989). Compliment-rejection versus compliment-avoidance: Listener-based versus speaker-based pragmatic strategies. *Language & Communication*.

Holmes, J. (1986). Compliments and compliment responses in New Zealand English. *Anthropological linguistics*, 485-508.

Holmes, J. (1988). Paying compliments: A sex-preferential politeness strategy. *Journal of pragmatics*, *12*(4), 445-465.

Jaworski, A. (1995). "This is not an empty compliment!" Polish compliments and the expression of solidarity 1. International *Journal of Applied Linguistics*, *5*(1), 63-94.

Leech, G. (1983). *Principles of pragmatics*. London and New York: Longman.

Lorenzo-Dus, N. (2001). Compliment responses among British and Spanish university students: A contrastive study. *Journal of Pragmatics*, *33*(1), 107-127.

Manes, J. (1983). Compliments: A mirror of cultural values. *Sociolinguistics and language acquisition*, *5*(3), 96-106.

Migdadi, F. H. (2003). *Complimenting in Jordanian Arabic: A socio-pragmatic analysis.*

Mustapha, A. S. (2004). *Gender variation in Nigerian English compliments* (Doctoral dissertation, University of Essex).

Nelson, G., Al-Batal, M., & Echols, E. (1996). Arabic and English compliment responses: Potential for pragmatic failure. *Applied Linguistics*, *17*(4), 411-432.

Pomerantz, A. (1978). Compliment responses: Notes on the co-operation of multiple constraints. In *Studies in the organization of conversational interaction* (pp. 79-112). Academic Press.

Pomerantz, A. (1978). Compliment responses: Notes on the co-operation of multiple constraints. In *Studies in the organization of conversational interaction* (pp. 79-112). Academic Press.

Pomerantz, A. (1984). *Agreeing and disagreeing with assessments: Some features of preferred/ dispreferred turn shaped.*

Ruhi, Ş. (2006). Politeness in compliment responses. *Pragmatics*. Quarterly Publication of the

International Pragmatics Association (IPrA), *16*(1), 43-101.

Ruhi, Ş., & Doğan, G. (2001). Relevance theory and compliments as phatic communication. *Linguistic Politeness Across Boundaries. The Case of Greek and Turkish*, 341-390.

Saito, H., & Beecken, M. (1997). An approach to instruction of pragmatic aspects: Implications of pragmatic transfer by American learners of Japanese. *The Modern Language Journal*, *81*(3), 363-377.

Scollon, R., & Scollon, S. W. (2001). Discourse and Intercultural Communication. *The Handbook of Discourse Analysis,* 538.

Sharifian, F. (2005). The Persian cultural schema of shekasteh-nafsi: A study of compliment responses in Persian and Anglo-Australian speakers. *Pragmatics & Cognition*, *13*(2), 337-361.

Spencer-Oatey, H., & Ng, P. (2001). Reconsidering Chinese modesty: Hong Kong and mainland Chinese evaluative judgements of compliment responses. *Journal of Asian Pacific Communication*, *11*(2), 181-201.

Tang, C. H., & Zhang, G. Q. (2009). A contrastive study of compliment responses among Australian English and Mandarin Chinese speakers. *Journal of Pragmatics*, *41*(2), 325-345.

Temple, B. (1997). Watch your tongue: Issues in translation and cross-cultural research. *Sociology*, 31(3), 607-618.

Wilson, D., & Sperber, D. (1993). Linguistic form and relevance. *Lingua*, *90*(1), 1-25.

Wolfson, N. (1983). An empirically based analysis of complimenting in American English. *Sociolinguistics and Language Acquisition*, 443, 82-95.

Yu, M. (2000). *Cross-cultural and interlanguage pragmatics: Developing communicative competence in a second language* (Taiwan).

Yu, M. C. (2003). On the universality of face: Evidence from Chinese compliment response behavior. *Journal of pragmatics*, *35*(10-11), 1679-1710.

Yu, M. C. (2004). Interlinguistic variation and similarity in second language speech act behavior. *The Modern Language Journal*, *88*(1), 102-119.

Yu, M. C. (2005). Sociolinguistic competence in the complimenting act of native Chinese and American English speakers: A mirror of cultural value. *Language and speech*, *48*(1), 91-119.

Yuan, Y. (2002). Compliments and compliment responses in Kunming Chinese. *Pragmatics*, *12*(2), 183-226.

NGHIÊN CỨU ẢNH HƯỞNG CỦA KHOẢNG CÁCH XÃ HỘI TỚI CÁC CHIẾN LƯỢC ĐÁP LẠI LỜI KHEN TRONG TIẾNG ANH MỸ VÀ TIẾNG VIỆT

Nguyễn Thị Thùy Linh

Trường Đại học Ngoại ngữ - ĐHQGHN,
Phạm Văn Đồng, Cầu Giấy, Hà Nội, Việt Nam

Tóm tắt: Nghiên cứu này được thực hiện nhằm mục đích điều tra ảnh hưởng của khoảng cách xã hội đối với việc lựa chọn chiến lược đáp lại lời khen trong tiếng Anh Mỹ và tiếng Việt. Để làm được điều này, hai bộ dữ liệu đã được thu thập bằng công cụ Bảng câu hỏi điền khuyết (DCT) với 12 câu hỏi tình huống trong đó người khen là những người có địa vị cao hơn, bằng hoặc thấp hơn nghiệm thể. Phân tích thống kê cung cấp những kết quả thống kê mô tả liên quan tới các chiến lược đáp lại lời khen ở 2 cấp đô, vĩ mô và vi mô. Nói cách khác, những kết quả này thể hiện các chiến lược tiếp nhận lời khen bao gồm: chấp nhận, sửa đổi, không chấp nhận, kết hợp và lảng tránh. Ngoài ra, số liệu thống kê suy luận còn chỉ ra rằng liệu có hay không một tiêu chuẩn chung trong việc lựa chọn các chiến lược đáp lại lời khen giống nhau giữa người Mỹ và người Việt. Cuối cùng, những kết quả thu được chỉ ra rằng có một ảnh hưởng nhất định giữa sự khác biệt trong khoảng cách xã hội và các cách thức đáp lại lời khen.

Từ khóa: khen, đáp lại lời khen, khoảng cách xã hội

APPENDIX

DISCOURSE COMPLETION TASK (DCT) QUESTIONNAIRE

Gender:……………………………………….

Thank you first for showing interest in participating in this study. You are kindly asked to fill out this questionnaire to contribute to a research project on "Compliment responses". What you are invited to do is to imagine yourself in a situation where you are being complimented by acquaintances and write down what you would say back to the compliments. There are a total number of 12 situations in this questionnaire.

In reply to the questions:

Please do this survey by yourself.

Please make the responses the way you think it is naturally occurring in real life situations.

If you would be prepared to take part in a follow-up group interview for the discussion of survey results, please let me contact via email or any kinds that you feel comfortable.

Contact: ……………………………………….

Thank you for your participation!

Situation 1.

You have just successfully defended your bachelor's thesis with high distinction. Your supervisor is really happy about that. She says, "*Well-done! It was a pleasure to work with you*!".

You say in response:

Situation 2.

You pay a visit to your parents-in-law at the weekend. Before the meal, you notice your mother in-law is busy preparing a big dinner. You approach and give her a helping hand without being asked. She really appreciates your enthusiasm saying, "*You're very thoughtful! It would have been a mess without your help*".

You say in response:

Situation 3.

After the new year holiday, you come back to work. To have a good start, you decided to pick the best outfit of yours in the wardrobe in the morning. On seeing you, your female boss says, *"You look great! I hardly recognize you today!"*.

You say in response:

Situation 4.

You have just had your first baby. Your boss and colleagues come to visit you and the baby at home. She says, *"Let me look at the little angel. What a cute chubby little boy!"*.

You say in response:

Situation 5.

You are a high school English language teacher. You have just had a lesson on the topic "How to write a persuasive opinion essay?". The students show a lot of interest in the lesson since it is rich in practical advice with valuable tips for essay writing. After class, a female student of yours comes to you and says, *"Thank you for your inspiring lesson! I learned a lot!"*.

You say in response:

Situation 6.

In the year-end party of your division, as the head manager, you have some nice words of appreciation for a hard-working year of your staff. Your speech is full of humors and your staff really enjoy it. A female subordinate comes to you afterwards and says, *"I didn't know that you're such a man of humor!"*.

You say in response:

Situation 7.

You have just lost some weight and look fitter after several weeks of intensive workout. Your little nephew notices this change. He says, *"Wow! You look like a movie star!"*.

You say in response:

Situation 8.

You are the director of a trading company. You have just bought a new sedan car and today you drive it to work. Your subordinate, seeing it comes to congratulate you on that. He says, "What a nice car!"

You say in response:

Situation 9.

You are a university student. With high scores, you are awarded a scholarship for the next semester. A male classmate of yours says, *"You deserve it! Way to go!"*.

You say in response:

Situation 10.

Your class is going on a field trip to a forest. Unfortunately, one of your female classmates trips over a stone and gets hurt. As the only boy nearby, you offer her a piggyback ride. She says, *"You're a great help! Thanks a lot!"*.

You say in response:

Situation 11.

You have just had your haircut today. Seeing you at the café, a male friend of yours says, *"Hey, you look 5 years younger!"*.

You say in response:

Situation 12.

After months of saving, you have managed to buy a new MacBook laptop. Your best friend is very happy for you. He says, *"You look smarter with this new MacBook!"*.

You say in response:

THE EFFECTIVENESS OF VSTEP.3-5 SPEAKING RATER TRAINING

Nguyen Thi Ngoc Quynh, Nguyen Thi Quynh Yen, Tran Thi Thu Hien, Nguyen Thi Phuong Thao, Bui Thien Sao*, Nguyen Thi Chi, Nguyen Quynh Hoa

VNU University of Languages and International Studies, Pham Van Dong, Cau Giay, Hanoi, Vietnam

Received 09 May 2020

Revised 10 July 2020; Accepted 15 July 2020

Abstract: Playing a vital role in assuring reliability of language performance assessment, rater training has been a topic of interest in research on large-scale testing. Similarly, in the context of VSTEP, the effectiveness of the rater training program has been of great concern. Thus, this research was conducted to investigate the impact of the VSTEP speaking rating scale training session in the rater training program provided by University of Languages and International Studies - Vietnam National University, Hanoi. Data were collected from 37 rater trainees of the program. Their ratings before and after the training session on the VSTEP.3-5 speaking rating scales were then compared. Particularly, dimensions of score reliability, criterion difficulty, rater severity, rater fit, rater bias, and score band separation were analyzed. Positive results were detected when the post-training ratings were shown to be more reliable, consistent, and distinguishable. Improvements were more noticeable for the score band separation and slighter in other aspects. Meaningful implications in terms of both future practices of rater training and rater training research methodology could be drawn from the study.

Keywords: rater training, speaking rating, speaking assessment, VSTEP, G theory, many-facet Rasch

1. Introduction

Rater training has been widely recognized as a way to assure the score reliability in language performance assessment, especially in large-scale examination (Luoma, 2004; Weigle, 1998). A large body of literature has been spent on how to conduct an efficacious rater training program and to what extent rater training program had impact on raters' ratings. More specifically, documents have shown that in line with general education measurement, rater training procedures in language assessment were also framed into four main approaches namely rater error training (RET), performance dimension training (PDT), frame-of-reference training (FORT), and behavioral observation training (BOT). The effectiveness of rater training and these approaches were the topic of interest for numerous researchers either in educational measurement or language assessment such as Linacre (1989), Weigle (1998), Roch and O'Sullivan (2003), Luoma (2004), Roch, Woehr, Mishra, and Kieszczynska (2011).

The same concern arose for the developers of the Vietnamese Standardized Test of English Proficiency (VSTEP). Officially introduced

* Corresponding author. Tel.: 84-968261056

Email: sao.buithien@gmail.com

in 2015 as a national high-stake test by the government, VSTEP level 3 to 5 (VSTEP.3-5) has been considered to be a significant innovation in language testing and assessment in Vietnam, responding to the demands of "creating a product or service with a global perspective in mind, while customising it to fit 'perfectly' in a local market" (Weir, 2020). This launching then led to an urgent demand of quality assurance in all processes of test development, test administration, and test rating. As a result, a ministerial decision on VSTEP speaking and writing rater training was issued in the later year (including regulations on curriculum framework, capacity of training institutions, trainer qualification and minimum language proficiency and teaching experience requirements of trainees). Being assigned as a training institution, University of Languages and International Studies (ULIS) has implemented the training program from then on. Inevitably, the impact of the rater training program has drawn attention from many stakeholders.

As an attempt to examine the effectiveness of the ULIS rater training program and enrich the literature of this field in Vietnam, a study was conducted by the researchers – also the organizer team of the program. In the scope of this study, the session on speaking rating scales, the heart of the training program for raters of speaking skill, was selected to investigate.

2. Literature review

With regard to performance assessment, there is a likelihood of inconsistency within and between raters (Bachman & Palmer, 1996; McNamara, 1996; Eckes, 2008; Weigle, 2002; Weir, 2005). Eckes (2008) synthesized various ways in which raters may differ: (a) in the degree to which they comply with the scoring rubric, (b) in the way they interpret criteria employed in operational scoring sessions, (c) in the degree of severity or leniency exhibited when scoring examinee performance, (d) in the understanding and use of rating scale categories, or (e) in the degree to which their ratings are consistent across examinees, scoring criteria, and performance tasks. (p.156). The attempt to minimize the divergence among raters was the rationale behind all the rater training programs of all fields.

Four rater training strategies or approaches have been described in many previous studies, namely rater error training (RET), performance dimension training (PDT), frame-of-reference training (FORT), and behavioral observation training (BOT). All of these strategies aim to enhance the rater quality, but each demonstrates different key features. While RET is used to caution raters of committing psychometric rating errors (e.g. leniency, central tendency, and halo effect), PDT and FORT focus on raters' cognitive processing of information by which the rating accuracy is guaranteed. Although PDT and FORT are similar in that they provide raters with the information about the performance dimensions being rated, the former just involves raters in co-creating and/or reviewing the rating scales whereas the latter provides standard examples corresponding to the described dimensions (Woehr & Huffcutt, 1994, p.190-192). In other words, through PDT raters accustom themselves to the descriptors of each assessment criterion in the rating scale, and through FORT raters have chances to visualize the rating criteria by means of analyzing the sample performances corresponding to specific band scores. The last common training strategy, BOT, focuses on raters' observation of behaviors rather than their evaluation of behavior. To put it another way, BOT is used to train raters to become skilled observers who are able to recognize or recall the performance aspects consistent with the rating scale (Woehr & Huffcutt, 1994, p.192).

A substantial amount of research in the field of testing and assessment has put an emphasis on rater training (Pulakos, 1986; Woehr & Huffcutt, 1994; Roch & O'Sullivan, 2003; Roch, Woehr, Mishra, & Kieszczynska, 2011; to name but a few) in an attempt for improving the rating, yet the findings about its efficiency seem to be inconsistently documented. Many researchers and scholars posited that RET reduced halo and leniency errors (Latham, Wexley, & Pursell, 1975; Smith, 1986; Hedge & Kavanagh, 1988; Rosales Sánchez, Díaz-Cabrera, & Hernández-Fernaud, 2019). These authors assumed that when raters are more aware of the rating errors they may commit, their ratings are likely to be more accurate. Nonetheless, the findings of Bernardin's and Pence's (1980) research showed that rater error training is an inappropriate approach to rater training and that this approach is likely to result in decreased rating accuracy. Hakel (1980) clarified that it would be more appropriate to term this approach as training about rating effects and that the rating effects represent not only errors but also true score variance. It means that "if these rating effects contain both error variance and true variance, training that reduces these effects not only reduces error variance, but affects true variance as well (cited in Hedge & Kavanagh, 1988, p.68).

In the meantime, certain evidence for the efficacy of rater training has been recorded for the other rating strategies, PDT (e.g. Hedge & Kavanagh, 1988; Woehr & Huffcutt, 1994), FORT (e.g. Hedge & Kavanagh, 1988; Noonan & Sulsky, 2001; Roch et al., 2011; Woehr & Huffcutt, 1994), and BOT (e.g. Bernardin & Walter, 1977; Latham, Wexley & Pursell, 1975; Thornton & Zorich, 1980, Noonan & Sulsky, 2001); particularly, FORT has been preferable for improving rater accuracy. However, Hedge and Kavanagh (1988) cautioned about the limited generalizability of the results in FORT. Specifically, in this training approach, the trainees are provided with the standard frame of reference as well as observation training on the correct behaviors. In other words, the results are dependent on the samples, which can hardly be generalized in all circumstances. Moreover, Noonan and Sulsky (2001) highlighted that FORT revealed weakness in that it did not facilitate raters in remembering specific test takers' behaviors, which might lead raters to false assessment in comparison to the described criteria.

In consideration of strengths and weaknesses of each training approach, an increasing number of researchers and scholars have had an idea of combining different approaches to enhance the effectiveness of rater training. For example, RET was combined with PDT or FORT (McIntyre, Smith, & Hassett, 1984; Pulakos, 1984), or FORT was combined with BOT (Noonan & Sulsky, 2001; Roch & O'Sullivan, 2003). Noticeably, no significant increase in rating accuracy has been reported. Nonetheless, the number of studies on the combination of different approaches was modest, which makes conclusion on its efficacy yet to be reached.

With a hope to enhance the impact on rating quality in the context of VSTEP, a combination of all four approaches was employed during the course of rater training program. However, similar to the general context with limited research on integrated approach in rater training, research in Vietnam has recorded to date few papers on language rater training and no papers on the program for VSTEP speaking raters, not to mention intensive training on rating scales. Therefore, it is significant to undertake the present study to examine whether the combination of multiple training strategies has an impact on performance ratings and what aspects of the ratings are impacted.

3. Research questions

Overall, this study was implemented to, firstly, shed light on the improvement (if any) of the reliability of the scores given by speaking raters after they received training on the VSTEP.3-5 speaking rating scales. Secondly, the study expanded to scrutinize the impact of the training session on other aspects namely criterion difficulty, rater severity, rater fit, rater bias, and score band separation. Accordingly, two research questions were formulated as follow.

1. How is the reliability of the VSTEP.3-5 speaking scores impacted after rater training session on rating scales?
2. How are the aspects of criterion difficulty, rater severity, rater fit, rater bias, and score band separation impacted after rater training session on rating scales?

4. Methodology

4.1. Participants

The research participants were 37 rater trainees of the rater training program delivered by ULIS. They worked as teachers of English carefully selected by their home institutions. Some prerequisite requirements for them to enroll in this course include C1 English proficiency level based on the Common European Framework of Reference (CEFR) or level 5 according to the CEFR – VN and at least 3 years of teaching experience. Additionally, good background on assessment is preferable. Some of them had certain experience with VSTEP as well as VSTEP rating, while the majority had the very first-hand experience to the test in the training course. With such a pool of participants, the study was expected to evaluate the rating accuracy of novice VSTEP trainee raters. It can be said that they were all motivated to take the intensive training program since they were commissioned to their study as the representatives of their home institutions, and some were financially bonded with their institutions. When being invited to participate in the study, all participants were truly devoted as they considered it a chance for them to see their progress in a short duration.

4.2. The speaking rater training program

A typical training program for speaking raters at ULIS lasts for 180 hours, consisting of both 75 hour online and 105 hour on-site training. The program is described in brief in this table below.

Table 1: Summary of rater training modules for speaking raters

	Theories of Testing and Assessment
Module 2	Rater Quality Assurance
Module 3	Theories of Speaking Assessment
Module 4	The CEFR
Module 5	CEFR Descriptors for Grammar & Vocabulary
Module 6	VSTEP Speaking Test Procedure
Module 7	VSTEP Speaking Rating Scales
Module 8	Rating practices with audio clips
Module 9	Rating practices with real test takers
Module 10	Assessment

As can be seen from the table, the training provided raters-to-be with both theoretical background and practical knowledge on VSTEP speaking rating. Even though trainees were experienced in their teaching and highly qualified in terms of English proficiency, testing and assessment appeared to be a gap in their knowledge. Therefore, the program firstly focused on an overview of language testing and assessment, then the assurance to maintain the quality of rating activity, followed by theories of speaking assessment as the key goal of this course. Due to the fact that VSTEP.3-5 is based on the CEFR, there was no doubt that there should contain some modules about this framework with an attention to three levels namely B1, B2, C1 as these levels are assessed by VSTEP.3-5. Moving on VSTEP's part, trainees were introduced to the speaking test format and test procedure. The rating scales would be analyzed in great detail together with sample audios for analysis and practice. The emphasis of the training program in this phase was for rating scale analysis and audio clip practice. The last practice activity was with real test takers before trainees were assessed with both audio clip rating and real test taker rating.

A spotlight in this training program is that it is designed as a combination of the four training approaches mentioned in the Literature review. To be more specific, in module 2, rater quality assurance, rater trainees were familiarized with rating errors that are generally frequent to rater, which demonstrated for the RET approach. Regarding module 4 and 5, when the CEFR was put into a detailed discussion, the FORT and PDT approach were applied. That is to say, the trainees' judgment on VSTEP's test takers was guided to align with the CEFR as a standardized framework to assess language users' levels of proficiency. From distinguishing "can-do" statements across levels in the CEFR, especially CEFR descriptors for Grammar and Vocabulary, trainees were expected to make some initial judgments of their future test takers using the CEFR as a framework of reference. In module 7 and 8, the application of all four approaches was clearly seen. At the beginning of the rating activity, rater trainees focused on the rating scales as the standard descriptions for three assessed levels known as B1, B2 and C1. Based on the level description of all criteria, trainees did their marking on the real audio clips of previous tests. Thus, this is a combination of both accustoming the raters-to-be to the descriptors of each assessment criterion as a signal of applying PDT and helping the trainees visualize the rating criteria by analyzing the sample performances with agreed scores from the expert rater committee as a signal of applying FORT. At the same time, RET was also used when trainees had a chance to reflect their rating after each activity to see if they make any frequent errors. Besides, BOT aiming at training raters to become skilled observer who are able to recognize or recall the performance aspects consistent with the rating scales was emphasized during all modules related to VSTEP rating activity. To illustrate, trainees were reminded to take notes during their rating, hence the notes help them link the test taker's performance with the description in the rubric. In this case, observation and note-taking did play a substantial role in the VSTEP speaking rating. The integration of mixed approaches in rater training, therefore, has been proved in this program.

4.3. Data collection

The data collection was conducted based on a pre- and post- training comparison. 37 trainees were asked to rate 5 audio clips of speaking performance before Module 7 where an in-depth analysis of the rating scales was performed. At this stage, they knew about the VSTEP.3-5 speaking test format and test procedure. They were also allowed to approach the rubric and work on the rubric on their own for a while. The 5-clip rating activity was conducted based on trainees' first understanding of the rating scales and their personal experience in speaking assessment. After a total of 20 hour on-site training in Module 7 and 8, the trainees involved in marking 10 clips including those 5 clips in random order. The reason why the initial 5 clips were embedded in the 10 later clips is that the participants were expected not to recognize the clips they had rated, which maintains the objectivity of the study. Rating the 10 clips is part of the practice session. The trainees' rating results were compared to those of an expert committee to check their accuracy. It is noteworthy to be aware that the clips used as research data were the recordings selected from practice interviews in previous training courses in which trainees were required to examine voluntary test takers. Both examiners and test takers in the interviews were anonymous, which guarantees the test security.

4.4. Data analysis

Multiple ways of analysis were exploited to examine the effectiveness of the rater training session. First of all, descriptive statistics of every rating criterion and of total scores were run. After that, traditional reliability analyses of exact and adjacent agreement, correlations, and Cronbach alpha were implemented. In order to further scrutinize the reliability, Generalizability theory was applied with the help of mGENOVA software. The approach of G theory utilized G study and D study to estimate variance component and dependability as well as generalizability of the speaking scores respectively. Finally, patterns of changes in rating quality were devolved into with many-facet Rasch analyses (FACETS software), in which how the criterion difficulty, rater severity, rater fit, rater bias, and score band separation was impacted after rater training session was unveiled.

5. Results

5.1. Descriptive statistics

Table 2: Descriptive statistics of speaking ratings before and after rater training
(37 trainee raters, 5 test-takers, 1 test)

	Grammar		Vocabulary		Pronunciation		Fluency		Discourse management		Total score	
	Mean	SD	Mean	SD	Mean	SD	Mean	SD	Mean	SD	Mean	SD
Pre-training	4.93	2.038	4.84	2.084	5.04	2.028	4.79	2.180	4.89	2.169	24.49	10.222
Post-training	4.98	2.040	4.94	2.240	4.96	2.055	4.91	2.198	4.97	2.308	24.76	10.599
Committee score	5.00	2.121	5.00	2.449	5.20	1.924	5.20	2.775	4.80	2.387	25.20	11.563

In the first place, mean ratings for each speaking criterion were presented in Table 2, which showed that the raters' scores were lower than committee scores in all criteria except for discourse management. Although the differences were modest the post-training scores of most criteria and the total score were closer to the committee scores than the pre-training ones. To investigate further into changes in ratings after training, analyses of traditional reliability were conducted.

5.2. Traditional reliability analyses

First of all, with the acceptable score difference set at 4 (out of 50 in total), the exact and adjacent agreement between raters and the committee was calculated. The result revealed that there were 148 scores within the acceptable range in total of 185 scores prior to the training session, which means the agreement rate was at 80%. The rate increased to 86% for the post-training ratings (159 out of 185 scores ± 4 points apart from the committee ones).

Besides exact and adjacent agreement, the inter-rater correlations were also computed. There were 666 significant inter-rater Pearson correlations resulted from 37 raters in total ($p<.05$). The average inter-rater correlation was high at .962 in the pre-training session and higher at .966 in the post-training session.

Finally, regarding Cronbach alpha index, the reliability level rose slightly from .986 to .988 after the raters received the training.

It can be seen that raters are already consistent before the training but there still existed improvement. As the changes were slight and seemingly negligible, more robust analysis methods were in need to scrutinize the patterns of improvement in aspects other than traditional reliability. This was the reason why G-theory and many-facet Rasch model were utilized.

5.3. Generalizability theory analyses

With the help of G study, variance components of the speaking scores were revealed in Table 3.

Table 3: Variance components of the speaking scores ($p^{\bullet}$ x $r^{\bullet}$ model, 5 test-takers, 37 raters)

	Variance source	Grammar		Vocabulary		Pronunciation		Fluency		Discourse management	
Pre	p (test-takers)	4.31029	86.25%	4.7006	89.28%	4.24264	85.80%	5.34775	92.20%	5.15488	90.17%
	r (raters)	0.07875	1.58%	0.10841	2.06%	0.1045	2.11%	0.12162	2.10%	0.13056	2.28%
	pr, error	0.60863	12.18%	0.45616	8.66%	0.5979	12.09%	0.33063	5.70%	0.43161	7.55%
	Total	4.99767	100%	5.26517	100%	4.94504	100%	5.8	100%	5.71705	100%
Post	p (test-takers)	4.51607	89.44%	5.59369	91.50%	4.70495	91.49%	5.46486	92.63%	6.10105	93.58%
	r (raters)	0.13559	2.69%	0.07297	1.19%	0.02628	0.51%	0.04835	0.82%	0.03018	0.46%
	pr, error	0.39745	7.87%	0.44685	7.31%	0.41126	8.00%	0.38649	6.55%	0.38814	5.95%
	Total	5.04911	100%	6.11351	100%	5.14249	100%	5.8997	100%	6.51937	100%

As indicated in Table 3, there are totally 3 variance components in G study conducted with $p^{\bullet}$ x $r^{\bullet}$ design: test-takers (p), raters (r) and the interaction between test-takers and raters.

Observably, approximately 85% to more than 90% of the speaking score variance were substantially from the test-takers, that is, the difference in the speaking scores is mainly caused by the disparity in students' proficiency levels. In contrast, the small percentage of variance coming from the main effect of rater variation source indicated that raters differed just slightly in their leniency/strictness and the difference was even narrowed from above 2% before training to negligible (1.19% or less) after training in four out of five criteria. In addition, it's noticeable that roughly 6% to 12% of the total variance in pre-training ratings was attributable to the variance component of the test-takers-raters interaction, which means the scores of test-takers varied to some extent across raters, especially for grammar and pronunciation. In the post-training round, this component explained less (6%-8%) of the total variance for all the criteria except for fluency. All these changes were the evidence for higher degree of consistency among raters after training.

Moreover, the D study also generated higher dependability and generalizability for the post-training ratings in all the criteria as well as the composite score (Table 4). Simply put, the ratings were more reliable after the training.

Table 4: Dependability and generalizability of the speaking scores ($p^{\bullet}$ x $r^{\bullet}$ model, 5 test-takers, 37 raters)

Criteria	Pre-training		Post-training	
	Φ (dependability)	Eρ² (generalizability)	Φ (dependability)	Eρ² (generalizability)
Grammar	0.99571	0.99620	0.99682	0.99763
Vocabulary	0.99676	0.99738	0.99749	0.99785
Pronunciation	0.99555	0.99621	0.99749	0.99764
Fluency	0.99772	0.99833	0.99785	0.99809
Discourse management	0.99706	0.99774	0.99815	0.99828
Composite (total score)	0.99778	0.99828	0.99858	0.99886

5.4. Many-facet Rasch analyses

Many-facet Rasch allowed the researchers to delve into the pattern of changes in different facets of the speaking test namely marking criteria, rater severity, rater misfit, rater bias, and band separation.

Regarding marking criteria, Figure 1 indicated that all five criteria of grammar, vocabulary, pronunciation, fluency, and discourse management gathered closely to each other on the difficulty scale, and even lined up after training. This enhancement emphasizes that all criteria were equally rated and no criterion was more difficult to fulfill than the others. The pronunciation criterion was shown to experience the most noticeable change when moving from the easiest position to the middle of the scale (Table 5).

When it comes to rater severity, Figure 1 showed that some raters became less severe and more raters clustered around the balanced point in the post-training ratings. This comes along with the decrease in the number of misfitting raters from eight to five out of 37 raters in total. These misfitting raters rated the test-takers' speaking performance differently from other raters, thus the infit mean square of their ratings was outside the desirable range from 0.5 to 1.5 (Table 6).

Table 5: Measures of the speaking criteria before and after training
(by ascending order of difficulty)

No.	Pre-training		Post-training	
	Criteria	Measure	Criteria	Measure
1 (easiest)	Pronunciation	-0.30	Grammar	-0.07
2	Grammar	-0.07	Discourse management	-0.04
3	Discourse management	0.01	Pronunciation	-0.03
4	Vocabulary	0.13	Vocabulary	0.04
5 (the most difficult)	Fluency	0.22	Fluency	0.10

Pre-training Post-training

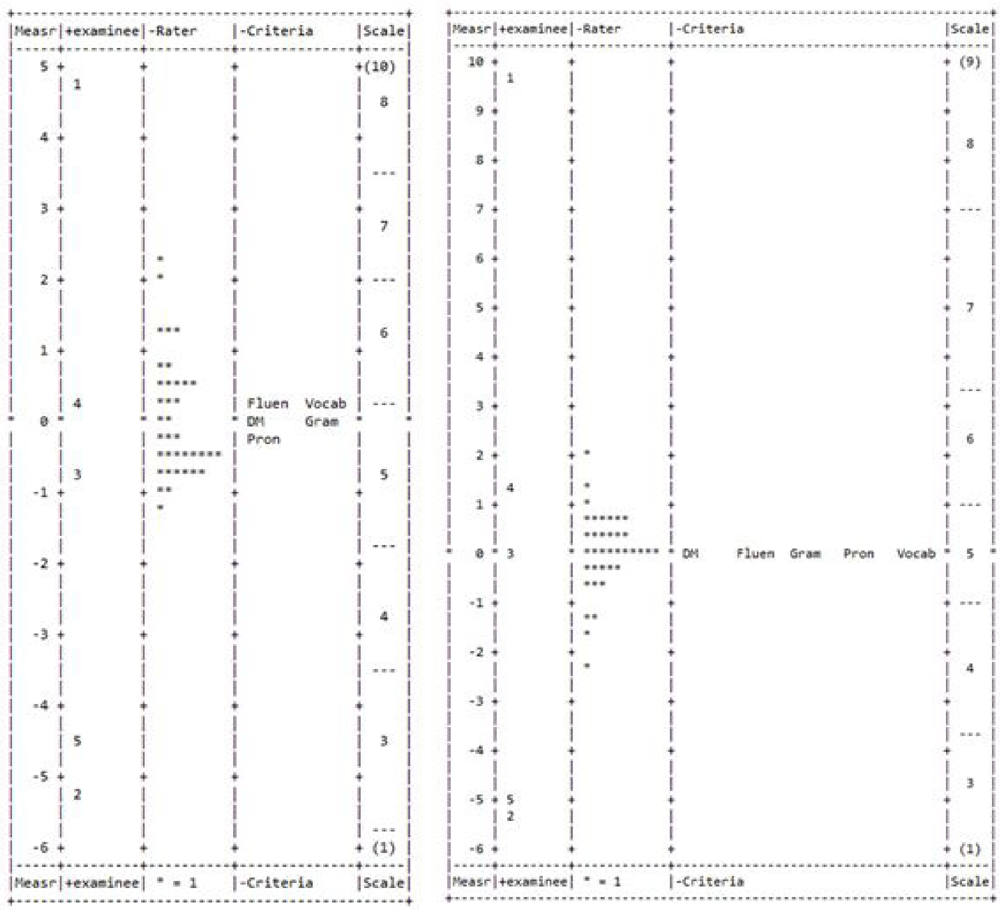


Figure 1: All facet summaries before and after training (5 test-takers, 37 raters)

Table 6: Rater fit indexes before and after training

Rater	Pre-training			Post-training		
	Measure	Infit		Measure	Infit	
		MnSq	ZStd		MnSq	ZStd
1	-1.16	0.67	-1.10	-0.03	0.70	-1.10
2	0.48	0.84	-0.40	1.11	0.68	-1.20
3	0.39	1.85	2.40	-0.13	1.43	1.50
4	-0.22	1.09	0.30	-0.03	0.93	-0.10
5	-0.65	0.92	-0.10	-0.13	1.06	0.20
6	1.17	0.53	-1.80	0.58	1.01	0.10
7	-0.57	1.10	0.40	-1.25	1.81	2.40
8	0.48	0.58	-1.50	0.07	1.05	0.20
9	-0.82	0.98	0.00	-0.33	0.89	-0.30
10	-0.22	1.03	0.10	-0.13	0.78	-0.70
11	-0.74	0.54	-1.70	-0.43	0.40	-2.80
12	-0.48	0.47	-2.10	-0.13	0.52	-2.00
13	-0.39	4.05	6.10	0.38	1.27	1.00
14	-0.39	0.64	-1.30	-0.23	0.94	-0.10
15	-0.57	0.40	-2.50	-0.03	0.63	-1.50
16	1.92	1.68	2.00	0.58	1.49	1.60
17	1.33	0.99	0.00	-0.33	0.71	-1.10
18	2.17	0.87	-0.30	-1.46	1.72	2.10
19	-0.48	1.34	1.10	-2.24	0.78	-0.60
20	0.65	0.65	-1.20	-1.67	0.63	-1.40
21	0.13	0.84	-0.40	2.17	0.69	-1.00
22	0.04	0.68	-1.10	0.58	0.63	-1.50
23	0.82	1.14	0.50	0.28	0.63	-1.50
24	0.56	0.65	-1.20	-0.53	0.99	0.00
25	1.33	1.42	1.30	-0.53	1.60	1.90
26	-0.99	0.54	-1.80	0.79	0.97	0.00
27	-0.82	0.97	0.00	-0.83	1.11	0.40
28	-0.13	1.43	1.30	-0.03	1.20	0.70
29	-0.57	1.24	0.80	0.38	0.65	-1.40
30	-0.57	0.46	-2.20	0.38	1.03	0.20
31	0.39	1.19	0.70	0.07	1.11	0.40
32	0.22	1.12	0.40	0.28	0.95	-0.10
33	-0.99	1.84	2.40	-0.23	1.08	0.30
34	-0.05	0.47	-2.10	0.18	1.82	2.50
35	0.30	0.96	0.00	0.69	0.70	-1.10
36	-0.82	0.55	-1.70	1.45	0.93	-0.10
37	-0.74	0.60	-1.40	0.69	0.58	-1.70

In addition, rater-occasion bias was also studied and 20 significant bias cases out of 74 bias terms were detected. Simply put, after training, ten out of 37 raters became significantly more lenient or severe than they had been in the pre-training phase.

Importantly, investigation in to score bands revealed better separation after training. Compared with Figure 2 which displayed that Band 4, 6, 7, and 8 were in overlap with other bands, Figure 3 showed a much more distinguishable separation for these central bands. Apparently, raters distinguish score bands considerably better after being trained. This is probably the biggest improvement in comparison with other aspects.

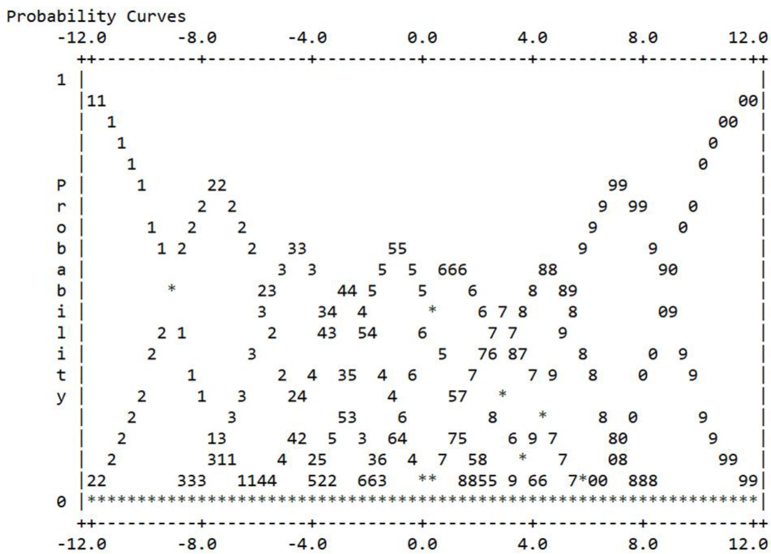


Figure 2: Score band separation before training

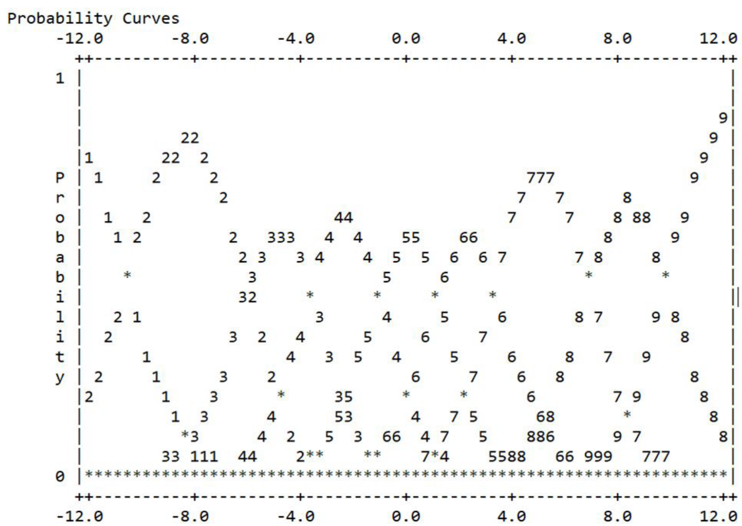


Figure 3: Score band separation afternoon training

6. Discussion

The study was implemented in order to find out the effectiveness of the rater training session on the VSTEP.3-5 rating scales. In general, the results have confirmed betterment in all the aspects examined.

With regard to the first research question, the reliability of the speaking scores analyzed both with traditional analyses and generalizability theory was showed to slightly increase after the training. In particular, higher values were recorded in exact and adjacent agreement among raters, inter-rater Pearson correlations, and Cronbach alpha reliability. The same is true for analysis with G-theory on the consistency among raters and the dependability and generalizability of the test scores.

Concerning the second research question, analyses with many-facet Rasch reported betterment in facets other than reliability: the balance in the difficulty of speaking criteria was enhanced, divergence in rater severity was lessened, rater fit was heightened, rater bias cases were fewer, and the score band separation was greater.

Generally, slight improvement was found for the majority of the aspects, and the most noticeable betterment was for the case of score band separation. Although the change was relatively small for some aspects of the speaking scores, it is still the evidence for the efficacy of the training session. Moreover, these positive changes can be considered important when taking other factors into consideration. Firstly, it is noteworthy in relation to the small number of training hours on the rating scales (20 hours). Furthermore, the pre-training rating session took place after Module 6, which means the raters already received a great amount of training on issues related to language assessment in general and speaking assessment and CEFR in particular. On top of that, research participants all had high-level qualifications and many years of experience in language teaching. All these factors likely helped the raters in shaping their ratings even before exposing to explicit guidance on the VSTEP.3-5 rating scale. Therefore, the impact is expected to be more visible and significant for either novice trainees or those yet to experience any training on standardized test scoring. This is also the researchers' suggestions for further research in the future.

Obviously, these above-presented results supported the point of researchers such as Smith (1986), Woehr and Huffcutt (1994), Noonan and Sulsky (2001), Roch et al. (2011), Rosales Sánchez, Díaz-Cabrera, and Hernández-Fernaud (2019), who had advocated and provided evidence for the enhancement of rating quality after rater training. Besides, the findings of small increase in score reliability was in line with several reports on slight improvement of rating accuracy by McIntyre, Smith, and Hassett (1984), Noonan and Sulsky (2001), Roch and O'Sullivan (2003). In addition to showing agreement with previous research, this study made meaningful contribution to literature in a way that it proved the effectiveness of a synthesized approach combining all four strategies of rater training and utilized various methods to statistically analyze the scores, both of which have not been widely documented in research so far. Moreover, it was this application of multiple statistical analyses that disclosed the noticeable enhancement in the score band separation of the speaking scores.

7. Conclusion

Overall, the study has rendered positive evidence for the efficacy of rater training, focusing on rating scale session with both guidance and practicing activities. After the training session, raters' ratings were found to be more reliable, consistent, and distinguishable.

Meaningful implications could be drawn from the study. Firstly, taking rater training administration into consideration, the combination of multiple/all training approaches is feasible and advisable. Although more research is needed to justify, the results suggest that if more combination is applied, greater impact is possible. This again not only restated the importance of rater training but also went beyond to emphasize the significance of how the rater training is conducted. Secondly, regarding methodological implication, the research showed that traditional statistical analyses through descriptives, Cronbach alpha, and correlations might not bring about sufficient information of the impact. In this case, the application of Generalizability theory and many-facet Rasch is recommended for better insights. Studies in the future should take this approach into consideration and expand to investigate the effectiveness of the whole rater training course for possible findings of significant changes.

References

Bachman, L. F., & Palmer, A. S. (1996). *Language testing in practice: Designing and developing useful language tests*. Oxford: Oxford University Press.

Bernardin, H. J., & Pence, E. C. (1980). Effects of rater training: Creating new response sets and decreasing accuracy. *Journal of Applied Psychology, 65*(1), 60–66. https://doi.org/10.1037/0021-9010.65.1.60

Bernardin, H. J., & Walter, C. S. (1977). Effects of rater training and diary-keeping on psychometric errors in ratings. *Journal of Applied Psychology*, *61*(1), 64-69.

Eckes, T. (2008). Rater types in writing performance assessments: A classification approach to rater variability. *Language Testing*, *25*(2), 155-185.

Hakel, M. D. (1980). *An appraisal of performance appraisal: Sniping with a shotgun*. Paper presented at the First Annual Scientist-Practitioner Conference in Industrial-Organizational Psychology, Virginia Beach, VA.

Hedge, J. W., & Kavanagh, M. J. (1988). Improving the accuracy of performance evaluations: Comparison of three methods of performance appraiser training. *Journal of Applied Psychology*, *73*(1), 68-73.

Latham, G. P., Wexley, K. N., & Pursell, E. D. (1975). Training managers to minimize rating errors in the observation of behavior. *Journal of Applied Psychology*, *60*(5), 550-555.

Linacre, J.M. (1989). *Many-faceted Rasch measurement*. Chicago, IL: MESA Press.

Luoma, S. (2004). *Assessing speaking*. Cambridge: Cambridge University Press.

McIntyre, R. M., Smith, D. E., & Hassett, C. E. (1984). Accuracy of performance ratings as affected by rater training and perceived purpose of rating. *Journal of Applied Psychology*, *69*(1), 147–156.

McNamara, T. F. (1996). *Measuring second language performance*. Essex: Addison Wesley Longman.

Noonan, L. E., & Sulsky, L. M. (2009). Impact of frame-of-reference and behavioral observation training on alternative training effectiveness criteria in a Canadian Military Sample. *Human Performance*, *14*(1), 3-26.

Pulakos, E. D. (1984). A comparison of rater training programs: Error training and accuracy training. *Journal of Applied Psychology*, *69*(4), 581-588.

Pulakos, E. D. (1986). The development of training programs to increase accuracy with different rating tasks. *Organizational behavior and human decision processes*, *38*, 76-91.

Roch, S. G., O'Sullivan, B. J. (2003). Frame of reference rater training issues: recall, time, and behavior observation training. *International Journal of Training and Development*, *7*(2), 93-107.

Roch, S. G., Woehr, D. J., Mishra, V., & Kieszczynska, U. (2011). Rater training revisited: An updated meta-analytic review of frame-of-reference training. *Journal of Occupational and Organizational Psychology*, *85*, 370-395.

Rosales Sánchez, C., Díaz-Cabrera, D., Hernández-Fernaud, E. (2019). Does effectiveness in performance appraisal improve with rater training? *PLoS ONE 14*(9): e0222694. https://doi.org/10.1371/journal.pone.0222694

Smith, D. E. (1986). Training programs for performance appraisal: A review. *Academy of Management Review*, *11*, 22-40.

Thornton, G. C., & Zorich, S. (1980). Training to improve observer accuracy. *Journal of Applied Psychology*, *65*(3), 351.

Weigle, S. C. (1998). Using FACETS to model rater training effects. *Language Testing*, *15*(2), 263–287

Weigle, S. C. (2002). *Assessing writing*. Cambridge: Cambridge University Press.

Weir, C. J. (2005). *Language testing and validation*. Hampshire: Palgrave McMillan.

Weir, C. J. (2020). Global, Local, or "Glocal": Alternative pathways in English language test provision. In L. I-W. Su, C. J. Weir, & J. R. W. Wu (Eds), *English Language Proficiency Testing in Asia: A New Paradigm Bridging Global and Local Contexts*. New York: Routledge.

Woehr, D. J., & Huffcutt, A. I. (1994). Rater training for performance appraisal: A quantitative review. *Journal of Occupational and Organizational Psychology*, *67*, 189-205.

HIỆU QUẢ CỦA HOẠT ĐỘNG TẬP HUẤN GIÁM KHẢO CHẤM NÓI VSTEP.3-5

Nguyễn Thị Ngọc Quỳnh, Nguyễn Thị Quỳnh Yến, Trần Thị Thu Hiền
Nguyễn Thị Phương Thảo, Bùi Thiện Sao, Nguyễn Thị Chi, Nguyễn Quỳnh Hoa

Trường Đại học Ngoại ngữ, Đại học Quốc gia Hà Nội
Phạm Văn Đồng, Cầu Giấy, Hà Nội, Việt Nam

Tóm tắt: Giữ vai trò quan trọng trong việc đảm bảo độ tin cậy của hoạt động kiểm tra đánh giá các kỹ năng sản sinh ngôn ngữ, tập huấn giám khảo (rater training) là một chủ đề thu hút trong nghiên cứu về các bài thi quy mô lớn. Tương tự, với bài thi VSTEP, hiệu quả của chương trình tập huấn giám khảo cũng nhận được nhiều sự quan tâm. Do đó, một nghiên cứu đã được tiến hành nhằm tìm hiểu ảnh hưởng của phần tập huấn sử dụng thang chấm Nói VSTEP.3-5 với các giám khảo trong chương trình bồi dưỡng tổ chức bởi Trường Đại học Ngoại ngữ - Đại học Quốc gia Hà Nội. Dữ liệu được thu thập từ 37 học viên tham gia khóa tập huấn nhằm so sánh việc chấm điểm của các học viên trước và sau phần tập huấn sử dụng thang chấm Nói. Cụ thể, các khía cạnh về độ tin cậy của điểm số, độ khó của tiêu chí, độ khó tính, độ phù hợp, và độ thiên lệch của giám khảo cũng như mức phân tách của thang điểm đã được phân tích. Nghiên cứu đã thu được các kết quả tích cực khi điểm số của các giám khảo đưa ra sau phần tập huấn có độ tin cậy, thống nhất, và phân tách tốt hơn. Sự cải thiện rõ rệt nhất được tìm thấy ở khía cạnh độ phân biệt mức điểm trong thang chấm. Một số ý nghĩa về hoạt động tập huấn giám khảo cũng như phương pháp nghiên cứu hoạt động này đã được rút ra từ các kết quả nghiên cứu.

Từ khóa: tập huấn giám khảo, chấm Nói, kiểm tra đánh giá kỹ năng Nói, VSTEP, lý thuyết G (G theory), phân tích Rasch nhiều khía cạnh (many-facet Rasch)

THE VALUE OF RATERS' COMMENTS ON THE WRITING COMPONENT OF A DIAGNOSTIC ASSESSMENT FOR LANGUAGE ADVISING

Stephanie Rummel*

University of Auckland,
Private Bag 92019, Victoria Street West, Auckland 1142, New Zealand

Received 15 March 2020
Revised 20 June 2020; Accepted 22 July 2020

Abstract: The Diagnostic English Language Needs Assessment (DELNA) is used at the University of Auckland to help identify the Academic English needs of students following admission in order to direct them to appropriate support (Elder & Von Randow, 2008). The second tier of DELNA is composed of listening, reading and writing sections, with the writing component rated by trained raters using an analytic rating scale. Language advisers then discuss the marking sheet with the student during an advisory session to provide a detailed overview of the strengths and weaknesses.

The current study was carried out because of difficulties language advisers were experiencing with utilising the marking sheets to draw students' attention to their strengths and weaknesses. A selection of 66 marking sheets with detailed comments from a variety of experienced raters was analysed and coded by two independent researchers. Themes were established regarding features that make a comment valuable or not valuable. Some of those same comments were then shared with students to determine whether or not they agreed with the advisers' assessment. The results show a mismatch at times between language advisers and students. The findings have been used to improve adviser practice and implement a more in-depth rater training programme to help raters better understand the descriptors and to utilise the rating scale to its full potential.

Keywords: Feedback, diagnostic feedback, feedback provision, feedback practices

1. Introduction

Universities in English-speaking countries are increasingly facing challenges as student populations become more linguistically diverse due to growth in the recruitment of international students, immigration inflows and initiatives to broaden participation in higher education by underrepresented groups (Read, 2016). In turn, a growing number of these institutions have begun to rely on post-entry diagnostic language assessments to identify students' academic language needs. According to Lee (2015), the purpose of diagnostics tests is twofold: to identify learners' strengths and weaknesses regarding specific elements of language use and to provide diagnostic feedback linked to remedial learning. These tests often assess students' academic reading, listening and writing skills with the intent of connecting students with resources that can help them appropriately develop in any areas where weaknesses have been identified. Procedures and processes vary among institutions, with the current study investigating the practices

* Tel.: +6493737599 ext 81844
Email: s.rummel@auckland.ac.nz; srummel444@yahoo.com

at the University of Auckland, with a specific focus on the value of comments provided by trained raters on the writing component of DELNA (Diagnostic English Language Needs Assessment), the institution's post-entry diagnostic assessment.

1.1. DELNA at the University of Auckland

DELNA is taken by all first-year students and PhD candidates and is a two-tiered assessment (Read & von Randow, 2016). Students first undertake a computer-based screening that takes about 30 minutes and includes a speedreading activity and an academic vocabulary task. The purpose of the screening is to provide an efficient way to identify proficient users of academic English and exempt them from further assessment (Read, 2008). However, if students fall under a pre-determined cut score, they are required to do a full two hour paper-based diagnosis (two and a half hours if they are a PhD candidate) of their listening, reading and writing skills.

Scores are reported on a scale ranging from 4-9 (Bright & von Randow, 2004). If students receive the highest bands, bands 8 and 9, it is unlikely that they will require academic English language support. Students receiving band 7 may benefit from some support, while band 6 students are thought to need concurrent academic English instruction. However, when a student falls into bands 4 or 5, they are considered at severe risk and in need of urgent language instruction. Those students then attend an advisory session and feedback is provided regarding their results.

1.2 The provision of feedback

According to Hattie and Timperley (2007), the definition of feedback is "information provided by an agent (e.g., teacher, peer, book, parent, self, experience) regarding aspects of one's performance or understanding" (p. 81). It has an important role in clarifying how well a person is doing and what needs improvement, which enables faster and more effective learning (Hounsell, 2003). Studies have identified various factors that make feedback either helpful or unhelpful. Maclellan (2001) claimed that students may improve their learning when they perceive the feedback to not simply be a judgement of their current level, but as a way to enable learning. Statements that are perceived as being judgemental or unmitigated statements have been found to be unhelpful or lead to defensiveness (Boud, 1995; Hounsell, 1995; Lea & Street, 2000). Weaver (2006) also found that students had difficulty understanding the feedback they received, with a main complaint being that it was too vague to be useful. A further issue identified by her participants was the need to balance negative comments with positive ones so that it would motivate students, which was also identified by Lee (2015) as being important in diagnostic assessments.

In order to be helpful, Lee (2015) posited that diagnostic feedback should establish links between various types of information. Furthermore, the feedback should not only reflect the diagnosis results, but also align itself closely with the resources and learning activities that are available (Lee, 2015). In order to facilitate this, different institutions have implemented varying procedures. Knoch (2012) found that academic advisors played a crucial role in conveying the results to students as they provide human contact in the process. In the case of DELNA, language advisers have delivered students' results since 2005. The position of language adviser was created in response to interview comments from students in which they expressed the desire to receive personalised advice during a one-on-one session (Bright & von Randow, 2004).

DELNA uses the diagnostic assessment to help students reflect on their strengths and weaknesses and a referral form to direct them to appropriate resources that promote academic language development. Any student who receives an average band of 6.5 or lower is asked to attend an advisory session with a DELNA Language Adviser lasting 30-40 minutes for non-PhD students. Any PhD candidate who undertakes the diagnosis attends a one-hour session regardless of their overall band. DELNA language advisers have backgrounds in academic English so they are well placed to help students interpret their results, with positive experiences being reported (Read & von Randow, 2016).

During the consultation, the adviser goes over a language profile that has been generated and includes overall band scores for the three skills that were assessed and computer-generated comments. Then the adviser focusses on the writing and, together with the student, reads through the comments provided by two trained raters regarding the student's writing. The original script is also consulted for specific examples that highlight the strengths and weaknesses. In this way weaknesses are "identified, represented, and described in a detailed and specific manner" (Lee, 2015, p. 304). Knoch (2011) argues that as much detail as possible should be provided from the results of a diagnostic assessment as detailed descriptions of the writer's behaviour allow with tips to improve future performances are more useful.

After various aspects of the writing have been carefully explained, the student is provided with information about workshops and online resources and given a referral sheet in both digital and hard copy to allow easy access. According to the original DELNA principles, there was to be an element of personal choice for students in that although they would be strongly recommended to take advantage of support, they should not be compelled against their will (Read, 2008). However, because questions have arisen regarding whether students actually follow up on recommendations when given the choice (Davies & Elder, 2005; Read, 2013; Knoch, Elder, & Hagan, 2016), currently participation in language enhancement options is required for students at the discretion of their academic programme (Read, 2013). This means that providing a clear description of students' strengths and weaknesses is important because some students may be required to show progress in their language skills before they can progress in their given programme.

1.3 DELNA rating

The quality of the rating is an important consideration in the interpretation of the results of any rater-mediated assessment (Hamp-Lyons, 2007; Johnson, Penny, & Gordon, 2009). In order to ensure validity and reliability, raters must be trained to use the scale to provide detailed feedback on student writing. Training is also important because rater variability may lead to issues such as construct-irrelevant variance (Barrett, 2001; Elder, Knoch, Barkhuizen, & von Randow, 2005; Weigle, 1998). Existing research has focused on rater reliability with issues such as the effectiveness of face to face and online rater training (Weigle, 1998) and rater bias (Weigle, 2011) being investigated, but these have all focussed on matching band scores.

The use of raters' marking sheets during the advisory session means that their comments play an important role in the feedback system utilised at DELNA. As such, on-going training is provided. Because the assessment is diagnostic in nature, it requires a different type of rating scale than those used for placement and performance, so an analytic scale has been chosen. According to Weigle (2002), analytic scales allow for an indication that different aspects of writing develop at different rates, which provides more useful diagnostic information. Currently, the scale includes nine traits clustered in three categories: coherence and academic style (text organisation, cohesion inside text and academic tone), content (description of data, reasons for trends observed, expansion of ideas), and form (sentence structure, grammatical accuracy, and vocabulary). Each trait is divided into six band levels ranging from four to nine. As raters rate, they are to fill out a marking sheet while referring to graded level descriptors for each trait. There is space on the marking sheet for raters to award a band for each of the nine traits, along with room for them to comment on each trait and provide ticks for correct uses of cohesive devices and referencing. They are also asked to provide crosses for incorrect uses of grammar and vocabulary and language impacting academic style, such as personal pronouns, contractions and informalities. It has been mentioned that some traits might not lend themselves to as fine distinctions as others, which could lead raters struggling to distinguish between the defined levels (North, 2003), so some traits may be more difficult to rate consistently than others.

Because raters' comments are shared with students, for DELNA it is vital that not only the scores match, but also the comments. Furthermore, the comments provide diagnostic information and language advisers must be able to use them to match students' needs with available support, but whether or not comments are valuable to language advisers and what makes a comment valuable have not previously been investigated. According to Kunnan and Jung (2009), "if diagnostic feedback provided to students is not dependable, its practical usefulness is cast into question" (p.617). DELNA language advisers have voiced issues with understanding and using some raters' comments in the past when providing feedback to students and directing them to resources, so the investigation of this issue seemed pertinent so that the training provided to raters could be improved.

2. Materials and Methods

2.1 Aims and research questions

This study aims to improve the comments provided by raters by examining the extent to which language advisers find the comments useful for advising students and students' perceptions of the comments. The research addressed the following questions:

1. What features make a rater's comment on a writing script for a diagnostic assessment valuable for a language adviser during an advisory session with a student?

2. What features reduce the diagnostic value of a rater's comment for a language adviser during an advisory session with a student?

3. To what extent do students' views of the usefulness of specific comments agree with those of the language advisers?

2.2 Methods

The research was carried out in two stages. In the first stage, which took place in 2017 and was used to answer research questions 1 and 2, a selection of 66 marking sheets with detailed comments from a variety of raters with a least two years of experience were chosen at random and analysed and coded by two independent researchers. One researcher was a current DELNA language adviser, while the other had previously been in the same position. Marking sheets were chosen at random to ensure there was a wide range of comments from different raters. It was decided that 66 sheets would provide a wide range of comments while at the same time allowing themes to emerge. Each marking sheet had raters' comments and band scores for three students on it and for each student there was to be one comment per trait for the nine traits. This means a total of 1,782 comments were analysed. The names of the raters on the rating sheets were covered to ensure anonymity so that the researchers would not be influenced by who had written the comments. The initial codes identified which comments were considered valuable by language advisers in that they allowed the advisers to provide constructive feedback related to specific aspects of students' writing such as grammatical forms, development of ideas, and academic style. The two researchers then worked together and further coding took place to establish themes regarding features such as specificity and clarity that made a comment either valuable or not valuable. This information was entered into a spreadsheet and themes were grouped together. The frequency of a comment being placed into a particular category was also tallied.

In the second stage, which took place in 2019, research question 3 was answered. An email was sent out inviting all students who had completed the diagnosis, received a band score of under 6.5, and been to see a Language Adviser in Semester 1. Five students contacted the DELNA office and all (n=5) were provided with a short survey that included some of the most frequently used comments and they were asked to comment on the usefulness of each. This was followed up with a one-on-one interview (n=4) to gain deeper insight into the students' perspective. Four students were English Language Learners (ELLs) from China, while one was a native speaker of English from New Zealand. One of the Chinese students was a PhD candidate. Of the four Chinese students, three were international students who had been in New Zealand for under a year and one was a permanent New Zealand resident who had been in the country for four years.

3. Results

3.1 Results for research questions 1 and 2

Types of comments that were considered valuable

A two-step process was used to first establish which comments were valuable or not valuable in their professional opinions. See Appendix A for a breakdown of each comment and its categorisation of usefulness. Please note that many comments were made more than once, so for the purpose of this report only each comment is recorded, not the number of times it was made. The researchers then worked together to establish what features made a comment valuable or not. For this step, comments were also checked against the other information on the marking sheets (band number and ticks and crosses) to identify any other issues that may have impacted the value of the comment.

A total of 83.73% (n=1492) of comments examined by the researchers were found to be valuable. The comments that were categorised as most valuable were clear and specific and closely mirrored the descriptors in the analytical scale. In those cases, it was very easy for the Language Adviser to understand why the rater had chosen the band, enabling the Adviser to direct students to appropriate resources. It was also helpful when raters provided information about both strengths and weaknesses that the student exhibited for a particular band. Examples of this were 'paragraphs exist, but topic sentences unclear' and 'splintered paragraphing, but some organisation of ideas'. The researchers found such comments provided both the Adviser and the student with valuable information about not only what they needed to improve, but also what they were doing well.

Consistency between the bands, the comments and the ticks/crosses was also valuable. It was helpful when the number given by the rater matched the comment provided, for example when a rater said there was some evidence of academic style, a phrase from the band 6 descriptor, and then in turn awarded band 6. In this case, Language Advisers could easily point out to students the areas where they needed improvement.

Another important point was that raters provided a clear comment for each of the nine categories. On the marking sheet, traits are given in the following order: (1) coherence, cohesion, and style; (2) content part 1, part 2, and part 3; (3) sentence structure, grammar, and vocabulary. It was helpful when raters commented in the order of the descriptors, making it clear which trait they were commenting on. Furthermore, when raters included examples in their comments, it was most valuable when they limited the number of examples provided to those that really highlighted the point they were making. Examples of informalities and correct and incorrect use of cohesive devices were particularly helpful because they were clear even when taken out of context.

Types of comments that were not considered valuable

The researchers found that 16.27% (n=290) of comments were not valuable (See Table 1 for specific details). The majority of issues centred around various inconsistencies with the raters' use of descriptor wording (n=145). The most common problem noticed by both researchers was that the comment matched a different band than the one given (n=102). One common example was related to academic style. To receive band 7, the descriptor states the writing should have "most aspects of academic style", for band 6 "some evidence of academic style" and for band 5 "little understanding of academic style". One rater commented that the writing showed "little sense of academic style", but then awarded band 6. At other times, the rater mixed wording from two or more descriptors or two or more traits. In one example, the rater gave band 8; however, the comment said "visible paragraphs, message clear, variable topics, shortish". The wording from this comment matches descriptors from bands 5 (shortish), 6 (variable topics), 7 (visible paragraphs), and 8 (message clear), so it was unclear why an 8 was given.

Other consistency issues were noted to a lesser degree. Raters sometimes double penalised students by, for example, marking them down in both style and vocabulary for informal language. There were also instances when raters penalised students in the wrong place. In the marking sheet there are three headings for comments: coherence/style, content, and form. An example of penalising students in the wrong place may be mentioning grammar errors under coherence/style rather than form and providing students with a lower band score as a result. Another issue arose when the ticks and crosses given by the rater did not match the comment (n=26). This issue was common in the form categories, where raters often commented that there were numerous grammar errors, but only provided one or two crosses across the categories.

Table 1: Categories of comments that were not valuable

Category	Frequency
Comment does not match band given	102
Examples listed with no context	29
Comment unclear/vague	48
Comment does not match ticks/crosses	26
No comment written	21
Mixed traits described in one comment	21
Comment under wrong trait	14
Difficult to read (handwriting, too much detail)	11
Harsh	10
Double penalisation	8

Both researchers found that some of the comments were unclear. In some cases, they simply did not make sense to the researchers (n=28). One such comment was "organisation is non-academic (has mixed parts)". Both researchers agreed that they were unclear as to what the rater meant. There were also times when the comments used very vague language (n=20) so the researchers were unable to discern the specific problem the rater had identified in the writing, for example "six paragraphs used".

Another issue impacting clarity was the quantity of information given. Some raters provided very detailed comments that became difficult to read given the limited amount of space provided. Others did not write comments for certain categories, often when ticks or crosses had been provided to show correct uses or errors. There were further cases when the raters simply provided lists of words as examples without context so the researchers could not decipher whether the students had used the examples correctly or incorrectly without consulting the original script.

The researchers found a few comments (n=10) that were not constructive as they seemed overly harsh or used too much jargon. Examples of this type of comment include "two topic sentences are non-sensical" and "reasons defy reason!"

3.2. Results for research question 3

In order to answer research question 3, student participants were provided with 17 comments that had been used often in the marking sheets that had been analysed in stage 1 to determine whether or not they found them useful. Most were comments that were found valuable by the language advisers, but a few were ones they thought were not valuable. Table 2 presents the comments language advisers found valuable and Table 3 presents those they felt were not valuable. Each table also includes how many students (n=5) agreed with the language advisers.

Table 2: Number of students who agreed with advisers that comments were valuable

Comment	Number of students who agreed (n=5)
Paragraphs exist, but topic sentences unclearParagraphs exist, but topic sentences unclear	4
Splintered paragraphing, but some organisation of ideas	2

Some paragraphs, but ideas lack organisation and there is repetition as well so it is hard to follow	4
Reasons are clear and well supported with logical development	5
Reasons are inadequate	3
Two reasons provided with adequate support	3
Good use of cohesive devices and clear referencing	5
Overuse of formulaic cohesive devices and repetitious referencing	4
Linking words used well to connect ideas	3
Occasional faulty reference	2
Inadequate range of vocabulary	5
A range of significant grammar errors	3
Article use requires attention	2

Table 3: Number of students who agreed with advisers that comments were not valuable

Comment	Number of students who agreed (n=5)
Organisation is non-academic (has mixed parts	2
Not quite visual paragraphs	3
Goes into substantial waffle about something off the topic	1
Walk/walked, their/there, are/was	2

Students were also asked to comment on why they found a comment valuable or not valuable. In general, when students found a comment to not be valuable, it was because they either did not understand it, or they wanted more specific information to help them understand it. For this reason, comments such as 'splintered paragraphing, but some organisation of ideas', 'occasional faulty reference', and 'article use requires attention' were found to be more valuable to language advisers than to students. The comment with the greatest difference was 'goes into substantial waffle about something off the topic'. Language advisers felt the comment was not valuable because it seemed a bit harsh and they worried that students would not know what was meant by 'waffle'. Students, however, found the comment to be valuable. When asked to explain what the comment meant, most focused on the second part of the comment, and understood they had written something unrelated. The native speaker of English understood the word 'waffle', and did not find it harsh. In the interview she said

> Um, I feel like a lot of lecturers mentioned the last point, about waffle, like don't feel as though you have to write a hundred pages 'cause it means you'll just waffle and completely miss the essay question, which is quite helpful for me…

Besides being given the comments, students were also asked in the interview whether seeing ticks and crosses was helpful. In response, the ELLs all felt it was helpful, with one stating "I think it will be better to get more specific example". However, the native speaker said: "It's not really nice seeing crosses, like what you didn't do. Um, more like maybe constructive feedback, like for next time do this…or you could have done this 'cause Xs can be quite off putting for some people."

4. Discussion

The findings from research question 1 and 2 of this study have implications for rater training in situations where raters are required to provide comments for feedback purposes. Because advisory sessions have been found to play a vital and helpful role in providing students with diagnostic information about their writing (Knoch, 2012; Schuh, 2008; Read & von Randow, 2016), it is important for raters to provide comments that the Language Advisers find useful. Traditional rater training often focuses on band scores; however, in instances when the assessment is diagnostic, comments are equally important as they can be used to better direct students to resources to work on identified weaknesses.

In the case of DELNA, the findings informed an expanded rater training programme for DELNA raters. In 2018, raters were provided with examples of valuable comments and comments that were not valuable and the trainer explained some of the factors that raters should consider when writing their comments. Emphasis was placed on the importance of writing comments that that were clear to they language advisers so that they could explain the comments to the students in language that would be accessible to them, even if they had low levels of language proficiency. Raters' attention was also drawn to key words in the different descriptors that highlight the differences between the bands, because the distinctions between them may not have previously been clear to raters (North, 2003). Furthermore, as most of the raters have experience as either teachers or IELTS examiners, the differences between the type of rating or grading they do in those situations and the type of feedback required for diagnostic assessments was also provided. After initial feedback from raters after the 2018 session, the 2019 training session was further expanded and returning raters were provided with some sample comments that were identified as not valuable and asked to categorise the comments under headings (for example: vague, harsh, etc). A discussion was also had regarding how the comments were used in the advisory session. It was hoped such activities raised raters' awareness so they have a better idea of how their comments are used and the ways they could be improved.

Some of the non-valuable comments were found in a limited number of marking sheets, suggesting they were provided by the same one or two raters. However, other issues such as a mismatch between the comment and the band were more universal. It would therefore seem pertinent to address those widespread problems in depth during the rater training with exercises that allow raters to become more familiar with the band descriptors. Issues that arose in only a few marking sheets could be mentioned during the training, but after rating begins if non-valuable comments are identified as coming from a specific rater, further feedback could be provided in an email.

Of all the identified issues, the frequency of raters awarding a band that did not match the comment is particularly worrying and has been brought to the raters' attention. Inter-rater reliability at DELNA is ensured by matching the marking sheets of two raters. However, only the band awarded is generally considered because there was an assumption that the band and the comment would match. In cases where the band and comment do not match, issues can arise during the advisory session if comments are conflicting, but have been given the same which information to provide to students, which can reduce the face validity of the assessment and also impact the advice being given.

Through raising raters' awareness and sharing experiences of when advisers meet students face to face, it is hoped that raters will give more thought to their comments. This is particularly true regarding the finding that comments that highlight both strengths and weaknesses are valuable, along with the findings that overly harsh comments are not helpful. Alderson and Huhta (2011) point out that diagnostic tests, due to their nature, have a greater focus on weaknesses than strengths. As such, most raters tend to focus on the negative aspects of the writing, but this may be demoralising for some students and that is not the purpose of the assessment. Because some faculties require students to complete a programme after meeting with the Language Adviser (Read, 2013), that they leave their session feeling positive and motivated to engage with the resources available to overcome their weaknesses in academic English is vital. Furthermore, according to Lee (2015), it is desirable to provide learners with information about their weaknesses in parallel with that of their strengths because, for an intervention on weaknesses to be successful, it needs to build on existing knowledge and skills that have already reached or neared the expected level. In this way, weaknesses and strengths may interact and impact the way a learner uses resources provided to enhance areas that have been identified as requiring improvement. The analytical feature of the DELNA scale was designed to allow for this because each criterion should be judged independently.

The findings have also started a discussion regarding the clarity of some of the items on the analytical scale and possible changes that may be made to the rating sheet. DELNA discussed the possibility of designing a rating sheet where raters highlight the relevant parts of the descriptors rather than write their own comments, which would eliminate issues with clarity, mixed descriptors and wrong choice of bands. However, there is a worry that important individualised diagnostic information could be lost if this decision is made. For that reason, it was decided to first provide more in-depth training regarding the comments to see if that would improve the results and raise raters' awareness.

Regarding research question 3, while there was agreement on the value of many, there was disagreement on others. Where there was disagreement, it was often because the student was unclear what the comment meant. This is why the language adviser role is important in the diagnostic feedback process. These comments were provided out of context; however, during the session, language advisers ask questions to try to ensure students understand. They also look through the student's script with them to point out specific examples related to the comments. Because the advisers are professionals in the field of academic writing, they are well placed to provide more explanation during the session and ensure students gain a better understanding of areas needing improvement.

The difference in the response of the native speaker to ticks and crosses is also interesting. As DELNA is administered to the entire student population, regardless of language background, it is important to be sensitive to how native speakers may view receiving feedback on their academic writing. They may also not be very aware of their weaknesses. DELNA seems to be slightly unique from other PELAs in that it is administered to the entire student population, regardless of language background. From experience, many ELLs enter the session with an awareness that their grammar and sentence structure may need some work, but often native speakers do not. Perhaps in those cases it is best to not focus so much on the crosses highlighting their errors and instead focus more on specific examples in the text that illustrate the point. This is already done in the language advising sessions, but by first showing some students the incorrect use of language in the form of crosses, they may be defensive before reviewing the script with the adviser. The same is true for comments that may be harsh. The goal during the session is to encourage students to use the resources available to improve any identified weaknesses, so it is important that it is not demotivating. However, it is difficult for language advisers to determine beforehand what students may deem as harsh, so language advisers need to be tuned in to students' responses and agile enough to make changes to the session so it suits each individual.

A limitation of the study is the small sample of student participants, so further recruitment could be done to provide a better representation of the student voice. Furthermore, the study could be expanded by investigating the issue from the raters' perspectives. Questionnaires or interviews with raters could be useful in determining reasons for the comments provided and allow for valuable information regarding raters' clarity surrounding the band descriptors. In addition, interviews or reflective journals from language advisers could provide better insight into reactions to the comments and the usefulness of various comments during advisory sessions.

5. Conclusions

The current study identified which comments provided by raters on a diagnostic writing assessment were deemed either valuable or not valuable. Although a robust body of research exists on rater reliability due to its impact on test validity and reliability, studies have mainly focused on test scores. The current study provides important insight into the type of rater training required when raters are asked to provide comments on the writing. In the past, the rater training focussed primarily on the band scores and ensuring raters had the same overall band; however, the findings from the current study emphasise the importance of providing raters with more guidance regarding comments when assessments are used for diagnostic purposes.

The language advisers are in a position to provide individualised feedback to each student who makes an appointment. The process is effective because they not only use the quantitative data contained in the score and the computer-generated comments provided on the profile, but also the qualitative data contained in raters' comments. When valuable comments are provided, they can enrich the advisory session and guide advisers to recommend appropriate resources for academic language enrichment; however, when the comments are not valuable, the adviser needs to spend extra time consulting the script and may even need to skip certain comments during the session. This is difficult during the busy period at the beginning of each semester when back to back appointments leave limited time for such preparation. The better understanding that raters have of how their comments are used and what is considered valuable, the better advisers can direct students. Therefore, enhanced training that goes beyond the band scores should lead to greater benefits for students.

References

Alderson, J.C., & Huhta, A. (2011). Can research into the diagnostic testing of reading in a second or foreign language contribute to SLA research? In L. Roberts, G. Pallotti and C. Bettoni (eds). *EUROSLA Yearbook 11*. John Benjamins, pp. 30-52.

Barrett, S. (2001). The impact of training on rater variability. *International Education Journal*, *2*(1), 49-58.

Boud, D. (1995). Assessment and learning: contradictory or complementary. *Assessment for learning in higher education*, 35-48.

Bright, C., & Von Randow, J. (2004). Tracking language test consequences: The student perspective. Paper presented at the IDP Australian International Education Conference, Sydney. Available online http://aiec.idp.com/uploads/pdf/thur%20-%20Bright%20&%20Randow.pdf

Davies, A. & C. Elder (2005). Validity and validation in language testing. In E. Hinkel (ed.), *Handbook of research on second language learning*. Mahwah, NJ: Erlbaum, 795–813.

Elder, C., Knoch, U., Barkhuizen, G., & von Randow, J. (2005). Individual feedback to enhance rater training: Does it work? *Language Assessment Quarterly: An International Journal*, *2*(3), 175-196.

Hamp-Lyons, L. (2007). Worrying about rating. *Assessing Writing*, *12*(1), 1–9. https://doi.org/10.1016/j.asw.2007.05.002

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

Hounsell, D. (2003). Student feedback, learning and development. *Higher education and the lifecourse*, 67-78.

Johnson, R. L., Penny, J. A., & Gordon, B. (2009). *Assessing performance: Designing, scoring, and validating performance tasks*. New York: The Guilford Press.

Knoch, U. (2011). Rating scales for diagnostic assessment of writing: What should they look like and where should the criteria come from?. *Assessing Writing*, *16*(2), 81-96.

Knoch, U. (2012). At the intersection of language assessment and academic advising: Communicating results of a large-scale diagnostic academic English writing assessment to students and other stakeholders. *Papers in Language Testing and Assessment*, *1*(1), 31-49.

Knoch, U., Elder, C., & O'Hagan, S. (2016). Examining the validity of a post-entry screening tool embedded in a specific policy context. In *Post-admission Language Assessment of University Students* (pp. 23-42). Springer International Publishing.

Kunnan, A. J., & Jang, E. E. (2009). Diagnostic feedback in language assessment. *The handbook of language teaching*, 610-627.

Lea, M., & Street, B. V. (2000). Student writing and staff feedback in higher education. *Student writing in higher education: New contexts*, 32-46.

Lee, Y. (2015). Future of diagnostic language assessment. *Language Testing, 32*(3), 295-298. doi:http://dx.doi.org.ezproxy.auckland.ac.nz/10.1177/0265532214565385

Maclellan, E. (2001). Assessment for learning: the differing perceptions of tutors and students. *Assessment & Evaluation in Higher Education*, *26*(4), 307-318.

North, B. (2003). Scales for rating language performance: Descriptive models, formulation styles, and presentation formats. *TOEFL Monograph*, 24.

Read, J. (2008). Identifying academic language needs through diagnostic assessment. *Journal of English for academic purposes*, *7*(3), 180-190.

Read, J. (2013). Issues in post-entry language assessment in English-medium universities. *Language Teaching, 48* (2), pp.1-18.

Read, J. (2016). Post-admission language assessment in universities: International perspectives. *Switzerland: Springer International Publishing*.

Read, J., & von Randow, J. (2013). A university post-entry English language assessment: Charting the changes. *IJES, International Journal of English Studies, 13*(2), 89-110.

Read, J., & von Randow, J. (2016). Extending Post-Entry Assessment to the Doctoral Level: New Challenges and Opportunities. In *Post-admission Language Assessment of University Students* (pp. 137-156). Springer, Cham.

Reinders, H. (2008). The what, why, and how of language advising. In: *MexTESOL*, *32*(2).

Schuh, J. H. (2008). Assessing student learning. In V. N. Gordon, W. R. Habley & T. J. Grites (Eds.), *Academic Advising: A comprehensive handbook*. San Francisco: Jossey-Boss.

Weaver, M. R. (2006). Do students value feedback? Student perceptions of tutors' written responses. *Assessment & Evaluation in Higher Education*, *31*(3), 379-394.

Weigle, S. C. (1998). Using FACETS to model rater training effects. *Language Testing*, *15*(2), 263-287.

Weigle, S.C. (2002). *Assessing writing*. Cambridge, UK: Cambridge University Press.

Weigle, S. C. (2011). Validation of automated scores of TOEFL iBT® tasks against nontest indicators of writing ability. *ETS Research Report Series*, *2011*(2).

GIÁ TRỊ NHỮNG NHẬN XÉT CỦA GIÁM KHẢO CHẤM VIẾT TRONG BÀI THI CHẨN ĐOÁN NHU CẦU TIẾNG ANH

Stephanie Rummel

Trường Đại học Auckland,
Private Bag 92019, Victoria Street West, Auckland 1142, New Zealand

Tóm tắt: Bài thi chẩn đoán nhu cầu tiếng Anh (DELNA) được sử dụng tại trường Đại học Auckland nhằm xác định nhu cầu về tiếng Anh học thuật của sinh viên sau khi nhập học; qua đó, bài thi sẽ giúp nhà trường cung cấp cho sinh viên những hỗ trợ phù hợp nhất (Elder & Von Randow, 2008). Bài thi DELNA hạng hai bao gồm kỹ năng nghe, đọc và viết. Trong đó, bài thi viết sẽ được các giám khảo chấm theo thang chấm phân tích. Các chuyên gia tư vấn ngôn ngữ sau đó sẽ thảo luận phiếu chấm cùng sinh viên trong các buổi tư vấn để mang tới cho sinh viên một cái nhìn tổng quan chi tiết về những điểm mạnh và điểm yếu của các em.

Nghiên cứu này được thực hiện khi các chuyên gia tư vấn ngôn ngữ gặp phải những khó khăn trong quá trình sử dụng phiếu chấm để làm việc cùng sinh viên. Nghiên cứu đã thu thập 66 phiếu chấm với những nhận xét chi tiết từ các giám khảo chấm viết dày dặn kinh nghiệm. Sau đó, hai nhà nghiên cứu độc lập đã tiến hành phân tích và mã hóa các phiếu chấm này. Nghiên cứu đã xác lập được các chủ đề liên quan đến những đặc điểm để đánh giá giá trị của một nhận xét. Một vài nhận xét giống nhau sau đó được gửi tới cho sinh viên để các em quyết định đồng ý hay không đồng ý với những đánh giá của các chuyên gia. Kết quả nghiên cứu cho thấy đôi khi có sự không đồng thuận giữa sinh viên và chuyên gia tư vấn. Những kết quả này đã được sử dụng để cải thiện hoạt động của các chuyên gia và tiến hành một chương trình đào tạo chuyên sâu hơn để giúp các giám khảo chấm viết hiểu rõ hơn về thang chấm và nhờ đó, sử dụng thang chấm hiệu quả nhất.

Từ khóa: phản hồi, phản hồi chẩn đoán, cung cấp phản hồi, hoạt động phản hồi

Appendix A: Raters' comments and whether they were valuable or not

Traits	Comment	Valuable	Not valuable
Coherence	Somewhat random paragraphing		✓
	Some organisation. No visual paragraphs	✓	
	Paragraphing clear. There is an introduction + topic sentences	✓	✓
	Two topic sentences are non-sensical		
	Paragraphing exists as do topic sentences. Message generally clear	✓	✓
	Organised in paragraphs but often needs re-reading		
	Visual paragraphs exist, but places content of some should be in others	✓	✓
	Reasons defy reason!		✓
	Paragraphs exist although a few too many. An introduction and a conclusion exist, but the former is a description, the latter is an irrelevance related to the internet in general	✓	✓
	Visual paragraphs present, but discussion poorly organised with data absent from part 1 but scattered across parts 2 and 3. No clear opening for Part 3	✓	✓
	Some paragraphs but ideas lack organisation and there is repetition as well. Hard to follow	✓	
	Includes some paragraphs but quite waffly and repetitive. Hard to follow. Possibly memorised		✓
	Includes paragraphs- message can generally be followed		✓
	Has used word to show introduction, but essay lacks paragraphs	✓	✓
	organisation is non academic (has mixed parts)		
	Paragraphs used for 3 parts, but few cohesive devices		✓
	Visible paras; messages clear; variable ts, shortish	✓	
	Opening/closing vague		✓
	Splintered paragraphs, short script. Breaks up part 2	✓	
	no visible paras; weak topics; some re reading		✓
	Introduction too general. Paragraphs used effectively to address parts of prompt	✓	
	Has paragraphs but they aren't esp helpful		✓
	Not quite visual paragraphs		
	Intro not very clearly developed/ ideas disconnected	✓	✓
	Ideas not always in logical order		
	Some organisation, some paragraphing. However some parts of the writing require rereading	✓	✓
	Lacks intro statement, only 2 paragraphs, poor org		
	Confused introduction		✓
	Inadequate introductory statement. Has 2 paras but p2 overly long, needs re-reading		
	Some reliance on rubric language		

Cohesion	Cohesion and reference are unnoticeable adequate CDs with some referencing Cohesive devices are seemingly simple or not quite accurately used Some incorrect CDs CDs: this, third, least most common, as or these due to, not just… but also SIMPLE Referencing: One main topic sentence found Some incorrect use of referencing. Formulaic simple cohesive devices are used. Some overuse of referencing such as they No referencing. Not many CDs Appropriate range of CDs with good referencing Some good use of CDs. Some used repeatedly. More referencing needed	✓ ✓ ✓ ✓ ✓ ✓	✓ ✓ ✓ ✓ ✓
Style	Chatty lexis interspersed with over formal phrases like ‘Proof of the above statement is shown or ‘can be obtained by’. Hedging adequate Style is appropriate although there is no hedging style is sometimes informal, and often simplistic. Hedging exists Many non-academic features: brackets for alternative grammatical structures; personal pronouns; chatty vocab Hedging exists Informality: the more we rely, rely too much mostly informality centres on direct address through pronouns Personal pronouns: who they Little understanding of academic style. Some p/p and rhetorical tone- “it tells us we should” Obscure/inconsistent logic Some evidence of academic style- some noticeable wordiness rhetoric- more and more Maintains formal register Formal but with many errors Little understanding of academic style. Too wordy/informal No actual problems apart from form Maintains formal tone + flow of logic. Prose not consistently intelligible. Wordiness Maintains academic distance but lacks analysis Some empty sentences Little understanding of academic style, spoken conversational lang, 1st person pronouns x 7, colloquialisms	✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓	✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓

Content	No NZ, or 2013, and little data, but overall statements are correct	✓	
	two trends mentioned but briefly		
	One +ve only is inferred		
	Comprehensive	✓	
	No NZ, no 2013, no figures, although trends accurate		
	data description includes place, but not time, and significant		✓
	figures and trends	✓	
	Interpretation is adequate and ideas are relevant with some		✓
	support		
	Time and place given as well as some significant	✓	
	figures(but one figure was misread or wrongly written		✓
	down) and no mention of figures for train or bicycle		
	Interpretation brief	✓	✓
	Ideas generally relevant		
	Interpretation is brief with some irrelevance	✓	✓
	Ideas generally relevant with some support		
	Interpretation is generally adequate and ideas are not	✓	✓
	always clear		
	Part 3 addressed	✓	✓
	Introduction is present, data and trends scattered through essay	✓	✓
	Paragraph 3 has content repeated from the middle of		
	second paragraph	✓	✓
	Mostly travel in cars…then walked vs. most		
	Some relevant ideas but they are not always relevant and		✓
	lack support	✓	
	Along with our health rate decreases		
	Goes into substantial waffle about something off topic	✓	
	Lacks trends but includes figures		✓
	Some reasons are based on assumptions that need	✓	
	substantiating and proof		
	Reason tangential- too much detail on an example	✓	✓
	Lacks overall trends, includes a run down of all figures		
	Some irrelevant reasons and assumptions	✓	✓
	Tangential answer-focused off topic		
	Description includes figures but lacks an overview		
	Lacks clarity- 2 figures 1 mode	✓	✓
	Some reasons for transport lack reason (catching a bus)		
	Gives place and year, notes data comes from a survey;	✓	✓
	gives main stats and trend		
	Combines trends with reasons, environment; price of bikes		✓
	and availability of bike racks	✓	
	Ideas not relevant enough		✓
	Convenience; proximity to work; more busses=fewer trains	✓	
	(x); not tightly structured		
	Partially described- general trends only		
	Very brief and inaccurate reasons	✓	
	Generally adequate		
	Facebook data ok, linkedIn not so detailed. Trends could be		
	more detailed		

Sentence Structure	Numerous fragments on page 1. Page 2, where there are	✓	✓
	complete sentences are grammatically accurate if unvaried	✓	✓
	Adequate if convoluted	✓	✓
	A variety of sentence types, mostly accurate. Punctuation errors	✓	✓
	A variety of sentence types, but rambling and some minor inaccuracies	✓	
	Some rambling and the continuous used for simple several times		
	Some convoluted sentences and punctuation errors		
	Frequent errors in sentence structures	✓	
	range of errors throughout such as punctuation, s+v agreement, sing/pl forms	✓	
	Punctuation: then walked or jogged, third most/then walked or jogged. Third most; , the least/The least	✓	
	Many incomplete sentences and wordiness which make script quite hard to read	✓	✓
	Complex forms contain errors- omissions or incomplete		
	Really wordy. Frequent errors in complex forms		✓
	Word order sometimes off, but most sentences are acceptable with just adequate range	✓	
	frag x1; a few awkward passages; most sentences correctly structured with some variety		
	Very sloppy sentence structure		
	Controlled and varied structures		
Grammar	Minor agreement and article errors	✓	✓
	Minor errors with verbs	✓	✓
	Some minor problems with grammar, especially sing/plural		✓
	Significant basic grammar errors and frequent vocab errors	✓	✓
	Limited control of sentence structure- incomplete and convoluted sentences	✓	✓
	Some basic grammar errors	✓	✓
	Some significant basic errors- articles and misplaced overuse of prepositions		
	On/in, too/to, go/went	✓	✓
	A few minor repeated errors- articles		
	Some repetition/sub + verb agreement? Collocations (higher parking fees)	✓	✓
	repetitious sub+verb agreement		
	People driving (no past tense) s+v agreement/word choice/ expression of ideas is awkward (buses trav. Ten or twentty minutes once) encourage people to catching bus		
	preps incorrect or missing; voice/tense errors; occasional missing word. Confuses be/do		
	missing article; possibly/le; tense; added 'which'		
	Walk/walked, their/there, are/was		

Vocabulary	Vocab accurate though lacks range	✓	✓
	Simple	✓	
	Vocabulary narrow and repetitive, and some oddities	✓	✓
	Vocabulary accurate but unvaried and a little imprecise	✓	
	Lexically unsophisticated	✓	
	A few wrong choices of vocab but generally appropriate		
	Range and use of vocab inadequate	✓	
	Many borderline vocab choices		
	Vocab is generally appropriate- limited range	✓	
	Range and use of vocab inappropriate- hard to understand	✓	
	Vocab adequate but not always sophisticated		✓
	A few spelling errors but generally appropriate vocab. Limited range	✓	
	Careful but shallow	✓	✓
	Some good vocabulary used, but limited range with grammar structures		

WHAT TEACHERS THINK ABOUT ASSESSMENT AND HOW IT INFLUENCES THEIR CLASSROOM-BASED ASSESSMENT PRACTICES – THE RESULTS OF A GLOBAL SURVEY

Susan Sheehan*

School of Education and Professional Development
University of Huddersfield, Queensgate, Huddersfield

Received 26 February 2020
Revised 16 May 2020; Accepted 24 July 2020

Abstract: This paper considers teachers' attitudes to assessment. The attitudes were explored through Borg's (2015) Language Teacher Cognition Framework. A mixed-methods approach of questionnaire and interview was adopted. The online survey was completed by English language teachers working in 57 different countries worldwide. The questionnaire was divided into 3 sections. The first section included questions which related to the participants' experiences of assessment at school. The second section explored the participants' assessment training experiences both in their initial teaching training and in any professional development sessions they had attended. The third and final section explored the participants' assessment practices and their beliefs about assessment. These three sections are based on the Language Teacher Cognition Framework. The interviews were conducted to explore the reasoning behind the responses given to the questionnaire. The framework also informed the data analysis process. Classroom experiences and professional development sessions were found to have the most influence on the teachers' attitudes to assessment. The participants were influenced by their experiences of assessment at school as they choose to avoid replicating the assessment practices which had been used when they were being assessed.

Keywords: Assessment, Teacher Cognition, Questionnaire

1. Introduction

This paper explores language teacher cognition and assessment. We investigated teachers' attitudes to assessment and how assessment influences their classroom-based assessment practices. We engaged with teachers of English as a foreign language (EFL) from around the world through an online questionnaire and follow-up interviews. The impetus for the project described in this paper was the puzzling results of a previous project (Berry, Sheehan and Munro, 2019) which had focussed on Language Assessment Literacy (LAL). We conducted a series of classroom observations which focussed on classroom-based assessment practices. We observed the teachers successfully deploying a range of assessment techniques in the classroom. When taking part in post-observation interviews those teachers all talked negatively about assessment and considered the observed assessment

* Tel.:44 1484 478144
Mail: S.Sheehan@hud.ac.uk.

activities to be part of their teaching practices. This led us to consider that the teachers had a fundamentally different understanding of assessment than the one we had developed through our reading of the LAL literature. We started to wonder if exploring the issue of assessment through the prism of teacher cognition would improve our understanding of classroom-based assessment practices

Much of the research into teachers' assessment practices has taken the approach of considering assessment to be a static body of knowledge which teachers have not mastered due to a lack of training in the topic or a lack of interest in the topic (Vogt and Tsagari, 2014). Our project, in contrast, explored assessment as a multi-layered and complex phenomenon which requires teachers to operate as both assessor and teacher (Scarino, 2013). Teachers have dual roles, and, at times, these roles can seem to be in conflict. For example, as an assessor a teacher may find a student's performance to be poor, but the teaching role could view the same performance as the starting point for a teaching intervention. It may be the case that the role of teacher is, perhaps unsurprisingly, more important to the teacher (Sheehan and Munro, 2017). Looney, Cumming, van Der Kleij, and Harris (2018) have also considered that teachers can have mixed feelings about assessment. These sentiments can be can be based on their feelings or experiences but also depend on their view of their role as teacher and if they are being asked or forced to engage with assessment activities which they feel are not part of their role as teacher.

Negative experiences of assessment have often been used as an explanation for poor levels of LAL. For example, Xu and Brown (2016) conducted an extensive review of studies which investigated assessment literacy. One of the conclusions which they drew from this review is that teachers' negative experiences of assessment impede their development of LAL. DeLuca, Chavez, Bellara and Cao (2013) are concerned that teachers have been adversely affected by their negative experiences of summative assessment. This leads them to view assessment negatively, but they also replicate the assessment types they were exposed to as they do not have alternative paradigms on which to draw. Ell, Hill and Grudnoff (2012) also raise the concern that teachers will reproduce the assessment practices they themselves experienced. Graham (2005) argues that teacher candidates in the US equated assessment with a test and that misapprehension was based on their experiences as students. This barrier to LAL has been encapsulated as teachers testing as they, themselves, were tested (Vogt and Tsagari, 2014).

Survey, to gather data about teachers' level of assessment literacy, has been a commonly used research strategy in both the fields of general education and English Language Teaching (ELT). The surveys, which have been conducted around the world with teachers working in a variety of educational settings, (e.g. Brookhart, 2011, Fulcher 2012, Hasselgreen, Carlsen and Helness. 2004, Mertler and Campbell, 2005) have generally presented teachers with items which relate to assessment and the teachers are asked if they are familiar with them. Some surveys further enquire if the teachers would like to receive training in those items with which the teachers felt themselves to be unfamiliar. Many of these studies have pointed to deficits in teacher knowledge about assessment. A limitation of survey research is that participants can only respond to the questions asked of them. The teachers may have been knowledgeable about other areas of assessment which had not been included in the surveys. Furthermore, the teachers who responded to surveys with requests for further training may have considered it as being professional to do so. The teachers were keen to expand all areas of their practice and the requests were an expression of interest rather than a cry for help.

As has been discussed above negative experiences of assessment have often been used as an explanation for low levels of LAL. In the literature relating to teacher education classroom experiences as children have been considered as having a strong influence on classroom practice. For example, Beijard, Meijer and Verloop (2004) suggest that teacher training is less influential on classroom practices than experiences at school. Thus, all childhood classroom experiences are influential on classroom practice and those which relate to assessment seem to be strongly influential. It may be the case the aspects of teacher training which relate to assessment do not influence classroom practices as childhood experiences of assessment are highly influential on classroom practice.

2. Materials and Methods

The following research questions (RQs) informed the project:

1. How do teachers develop their identity as assessors?
2. What role do teachers experiences of assessment, both in their own schooling, and as teachers, play in the development of their assessment practices?
3. How do teachers put their assessment 'credo' into practice?

The three research questions were developed following completion of the literature review. RQ1 was developed from Looney et al's (2018) model of Teacher Assessment Identity (TAI). Looney et al (2018: 14) state: "… we propose a dynamic and interactive assessment identity constituted by beliefs, feelings, knowledge and skills." RQ2 reflects Borg's (2015) framework for language teacher cognition. This framework will be discussed in Section 3 of this paper. The term credo in the third research question was inspired by Yin's (2010) study which looked at how teachers made assessment related decisions both in the classroom and when writing lesson plans. He argued that teachers had a set of beliefs or 'credo' which underpinned all decisions and practice. The data presented in this paper mainly refer to RQ2. Further details of the whole project and the data collected for the first and third RQ can be found in Sheehan and Munro (2019).

Mixed-methods strategy.

A mixed-methods approach (Cohen, Manion, Morrison, & Bell, 2011; 2013) was adopted. Data were collected through questionnaire and follow-up interview. The questionnaire had been inspired by Borg and Burns' (2008) survey of teacher beliefs about teaching grammar. The interviews were conducted to explore the reasoning behind the responses given to the questionnaire. The questionnaire was divided into 3 sections. The first section included questions which related to the participants' experiences of assessment at school. The second section explored the participants' assessment training experiences both in their initial teaching training and in any professional development sessions they had attended. The third and final section explored the participants' assessment practices and their beliefs about assessment. The questionnaire started with biographical questions such as years of teaching experience, qualifications and the type of institution in which they taught most often. We recognise that many teachers work for a variety of different institutions both in the public and private sectors. Following Borg and Burns (2008) we did not ask the participants if they considered themselves to be own language speakers of English or not as these distinctions seem to be reductive. The questions in the main section of the questionnaire were 5-point Likert scale. There was one open response question to allow participants to make any comment they wanted to. As we planned to conduct interviews, we did not include open response items on the questionnaire. At the end of the questionnaire we asked participants who were willing to be interviewed to leave their email address. The online questionnaire was promoted by the British Council and through the professional networks of the researchers. The interviews were conducted using Zoom. The semi-structured interviews were also based on the three identified aspects of Borg's (2015) framework.

Below is a table which summarises the participant numbers and their locations.

Table 1. Participant numbers and their locations

Data Collection Stage	Participants	Location
Questionnaire	261 participants	Worldwide, with participants from 57 different countries
Interviews	10 interviews	Worldwide

We used the following definition of teacher by Vogt and Tsagari: "Someone who is a practising EFL teacher who has undergone regular training to teach English as a foreign language at state or private tertiary institutions, colleges or schools." (2014:377). Our participants taught all ages and levels of students from young learners to adults, including general English and English for Specific Purposes. Responses were obtained from five different continents. 240 participants stated they worked in countries where English was taught as a foreign language. The remaining 21 participants worked in countries where English is taught as second language. Of this group the majority (17) were working in the UK. There was great variation in the language policies of the countries where the participants worked. One of the limitations of conducting research on a global scale is that the opportunities for exploring national language policies are, necessarily, limited. Some countries and regions were more represented such as India and South America which may reflect British Council activity in those countries. 32 respondents were from Vietnam. The sample was essentially self-selecting. The use of the word assessment in the questionnaire title may have been off-putting for some teachers so the respondents could be those teachers with a special interest in assessment. The participants could be described as being experienced as 51.3% had 15 years plus experience as a teacher. They could also be described as well-qualified as 34.1% had master's qualifications and 7.3% held a doctorate. This final figure may reflect use of the International Association of Teachers of English as a Foreign Language (IATEFL) research special interest group as one of the promotion channels for the questionnaire. A special interest group which focusses on research might be expected to have a number of members who hold doctorates. 48.7% of the participants stated they worked in state sector most often. It may well be the case that several participants worked in a variety of different sectors. Secondary/high school was the most common workplace with 32.2% of participants stating they worked most often in this type of institution.

The questionnaire data were analysed using descriptive statistics. The interview data were analysed thematically. The themes were developed from Borg's (2015) Language Teacher Cognition Framework. This framework is discussed below.

3. Theory

Language Teacher Cognition

Borg (2015:321) defines language teacher cognition as: "… an inclusive term referring to the complex practically-orientated, personalized, and context-sensitive networks of knowledge, thoughts and beliefs that language teachers draw on in their work." This definition highlights the complexity of the proposition and its multi-faceted nature. This framework was chosen because of the focus on language teachers. Borg (2015) highlights the ways in which language teaching differs from the teaching of subjects which means that other teacher cognition frameworks would not have been appropriate for this project. In addition, the framework considers teacher cognition to be context-sensitive. This is of special relevance to a project which focuses on assessment as the assessment policies of country, for example, the requirement to pass an English examination as part of college entry requirements can influence the classroom-based assessment practices.

The project drew on Borg's Language Teacher Cognition Framework. The framework includes: Schooling, Professional Coursework and Classroom Practices. These factors impact on language teacher cognition. The relationship between these factors and language teacher cognition is two-way. RQ2 was based on this framework.

4. Results

A key and unexpected finding was that 74.4% of participants stated they did not use the same techniques as their teachers used with them. The significance of this result will be discussed below. One of the interviewed teachers stated: "At school we never knew what was expected from us…the grades were shady you never knew why you got a three or a four." This person, and she was one among several, lamented the opaqueness of the marking system used at her school. She did not understand what she needed to do to gain high grades and she was not clear about the rationale behind the scores she was awarded. 51.3% of participants stated school assessment experiences influenced assessment practices. This could seem to contradict the previously stated finding. A review of the qualitative data may suggest that the two different findings are not, in fact, contradictory. The following interview quotations, selected from many which expressed similar views, may help to understand the situation better. One participant stated: "My experience taught me the things I should stop doing." Another participant said: "We just had tests we didn't get feedback." So, there is evidence to suggest that participants were actively seeking to avoid repeating the assessment practices which they had experienced as school children.

Few participants stated they had experienced assessment activities associated with assessment for learning such as self- or peer-assessment as children. The participants stated they used such activities regularly. For example, 66.8% of the questionnaire respondents stated they used self-assessment. One of interviewed teachers explained his use of peer assessment thus: "I use peer assessment all the time, the power relationship is more symmetrical, and they are more inclined talk to each other so that lowers the filter and all that. It encourages negotiation and clarification, encourages noticing and critical thinking, and it encourages a more student-centred approach." Nearly half (48.5%) of the questionnaire respondents stated that they had not experienced self-assessment or similar activities when they were students themselves but stated that they used them as teachers. This is further support for the claim that the participants are not replicating the assessment practices which they themselves had experienced. It should also be noted that contextual factors can limit the range of assessment activities used in the classroom. One participant in Thailand stated: "I do my best to work with assessment for learning, they find it strange because the culture here is having tests."

90.2% of participants stated they monitored students throughout the lesson. This could be considered an unsurprising finding. Assessment and teaching are inter-twined or interwoven. One participant stated that it was classroom experience which helped her to understand the importance of assessment. She stated: "Everything starts with assessment – it took me a long time to learn this – we start thinking with good achievement and this is how assessment is part of planning, assessment is part of teaching, so in thinking about how I'm going to assess I decided how to teach and plan a topic, how it will be assessed throughout the lessons…I cannot think about teaching without thinking about assessment." These findings suggest that classroom experiences as a teacher are highly influential on assessment practice and that assessment practices develop out of reflections on classroom practice. This further quotation exemplifies the influence of teaching practice on the development of assessment practices. One teacher stated: "At the beginning I reproduced what I was exposed to. It was only when I started having to explore different ideas I realised I could do something different."

While some participants mentioned experience as a key influence on the development of their assessment practices others highlighted the importance for them of attending Continuous Professional Development (CPD) courses. For example, a participant from Poland stated that a training course delivered by the Peace Corps had led her to change her assessment practices: "So I realised it wasn't the way to motivate students. I realised I was closing the loop." She completely changed her assessment practices after attending a training course on assessment for learning. She went on to explain that when she had used pen-and-paper tests with the students the dialogue stopped when the students had received their mark. When she used the assessment activities, which she had learnt on the course she was able to create more constructive dialogues with her students which developed over the whole academic year. Another participant stated: "I got to learn about formative assessment, so I started changing my way of assessment."

A further example of the importance of CPD course on the development of assessment practices relates to the introduction of Common European Framework of Reference (CEFR). Several participants indicated they had attended courses on the CEFR. One participant from Brazil stated: "I did a course on the CEFR – they encouraged the use of the portfolio, for self-evaluation and for us to evaluate them." This would seem to suggest the action-orientated approach to language learning and assessing as described in the CEFR has had an influence on classroom-based assessment practices through CPD courses.

Contextual factors were also influential on classroom-based assessment practices. In the interviews a number of teachers mentioned the difficulties of meeting parental expectations. A teacher in India stated that the parents demanded that she gave her students a mock test every week. Such a practice did not accord with her beliefs on assessment but she felt she was not in a position to refuse the request. This comment was made by a single teacher but illuminates the pressures some teachers are under. She stated: "Authorities can come and check your books or tests without warning – a colleague did receive criticism from the authorities." It would seem that some teachers are subject to severe scrutiny and face a heavy burden of accountability. Contextual factors should also be taken into consideration when reviewing the results of this project. The teachers who responded to the survey may be more proactive or they may have more favourable working conditions with associated provision of training and development opportunities. Thus, their experiences may not reflect those of teachers who are not able to engage with the British Council or who do not have the opportunity to participate in research projects.

5. Discussion

We had anticipated that assessment experiences as a language learner would be influential on teachers' assessment practices with teachers testing as they had been tested (Vogt and Tsagari, 2014, Xu and Brown, 2016, Ell, Hill and Grudnoff, 2012). However, the participant teachers made a conscious decision not to replicate these practices as they were aware of the shortcomings of pen-and-paper tests even though they had tended to score high marks on such tests. There is evidence to suggest that participants were actively seeking to avoid repeating the testing practices which they had experienced as school children.

The participants emphasised the importance of self- and peer-assessment activities in their classroom-based assessment activities. These findings echo those reported in our previous study of teacher assessment practices (Sheehan and Munro, 2017). The teachers stated that such practices promoted a more holistic approach to language learning which encouraged leaners to focus on developing their level of language proficiency and not on test scores. This finding is in contrast with that of Graham (2005) who found that the participants in her study had a rather limited conception of assessments and tended to confuse tests for assessment. Few participants reported they had experienced self-assessment as students, so classroom teaching experience would seem to be highly influential on language teacher cognition.

Experiences in the classroom and participation in continuous professional development courses were found to be more influential in the development of teachers' assessment practices and beliefs that either their schooling or initial training qualifications. This finding would seem to suggest that Beijard et al's (2004) proposition that initial teacher training is not influential on classroom teaching practice is accurate. This project, however, did not find evidence to support the notion childhood classroom experiences were influential on teaching practice. Rather, this study found that CPD and time spent teaching in the classroom had the largest influence on teachers' classroom-based assessment practices.

The use of Borg's language teacher cognition framework allowed us to explore the teachers' attitudes to assessment and how these influenced their classroom-based assessment practices in a nuanced way. The different parts of the framework and the bi-directional relationship between them helped us to tease out the different influences on classroom practice. The findings of this study are also in accord with Scarino's (2013) view of assessment as a multi-layered and complex phenomenon. In addition, the use of framework allowed us to explore the influence of contextual factors on assessment practices. These factors included parental pressures and on-the-spot inspections of teaching materials and tests.

This project, like all projects, has limitations which should be acknowledged. The sample was self-selecting and it may have attracted teachers with a particular interest in assessment. The survey was only available online so teachers without reliable internet access were excluded from the study. The questionnaire was promoted through a number of different channels such as British Council websites and Twitter. Teachers who do not engage with such channels were also excluded from the study. The project is based on self-report data and this might not accurately reflect classroom practice. It is a representation of what the teachers say they do in the classroom. The survey did attract respondents from around world so in one sense it can be considered to be a global survey. The number of participants, however, represents only a tiny fraction of the total number of EFL teachers in the world.

6. Conclusions

In sum, the teachers in this project do not appear to 'test as they were tested' (Vogt and Tsagari, 2014). Teaching experience and engagement with CPD activities are both highly influential on the development of classroom-based assessment practices. The teachers promoted a holistic and student-centred approach to language teaching and learning. Assessment is a tool to support this approach. We found that teacher cognitions are most influenced by teaching experiences and training courses.

In terms of teacher education, we would make three recommendations. Firstly, teachers should be encouraged to reflect on their own experiences of assessment and reflect on how these have influenced their assessment practice. Secondly, teacher-training courses should focus more on classroom-based assessment activities. Lastly, trainee-teachers' assessment practices should be discussed during feedback on teaching practice sessions.

References

Beijaard, D., Meijer, P. C., & Verloop, N. (2004). Reconsidering research on teachers' professional identity. *Teaching and Teacher Education, 20*(2), 107-128. doi:10.1016/j.tate.2003.07.001

Berry, V., Sheehan, S., & Munro, S. (2019). What does language assessment literacy mean to teachers? *ELT Journal, 73*(2), 113-123. doi:10.1093/elt/ccy05

Borg, S. (2006) *Teacher cognition and language education*: Research and practice. London: Continuum.

Borg, S. (2015). *Teacher cognition and language education: Research and practice* (1st ed.). London, England; New York: Bloomsbury Academic.

Borg, S and Burns, A. (2008). Integrating grammar in adult TESOL classrooms. *Applied Linguistics 29*(3), 456-482

Brookhart, S. M. (2011). Educational assessment knowledge and skills for teachers. *Educational Measurement: Issues and Practice, 30*(1), 3-12. doi:10.1111/j.1745-3992.2010.00195.x

Cohen, L., Manion, L., Morrison, K., & Bell, R. (2011; 2013). *Research methods in education* (7th; Seventh; ed.). London; New York; Routledge. doi:10.4324/9780203720967

DeLuca, C., Chavez, T., Bellara, A., & Cao, C. (2013). Pedagogies for preservice assessment education: Supporting teacher candidates' assessment literacy development. *The Teacher Educator, 48*(2), 128.

Ell, F., Hill, M., & Grudnoff, L. (2012). Finding out more about teacher candidates' prior knowledge: Implications for teacher educators. *Asia - Pacific Journal of Teacher Education, 40*(1), 55.

Fulcher, G. (2012). Assessment Literacy for the Language Classroom. *Language Assessment Quarterly 9*(2), 113-132. doi:10.1080/15434303.2011.642041

Graham, P. (2005). Classroom-based assessment: Changing knowledge and practice through preservice teacher education. *Teaching and Teacher Education, 21*(6), 607-621. doi:10.1016/j.tate.2005.05.001

Hasselgreen, A., Carlsen, C. and Helness, H. (2004). *European Survey of Language Testing and Assessment Needs. General Findings*, On WWW at http://www.ealta.eu.org/resources.htm.

Looney, A, Cumming, J, van Der Kleij, F and Harris, K. (2018). Reconceptualising the role of teachers as assessors: teacher assessment identity. *Assessment in Education: Principles, Policy & Practice*, *25*(5), 442-467.

Mertler, C. A., Campbell, C. S. (2005, April). Measuring teachers' knowledge and application of classroom assessment concepts: Development of the Assessment Literacy Inventory. Paper presented at the annual meeting of the American Educational Research Association, Montreal, Quebec, Canada.

Scarino, A. (2013). Language assessment literacy as self-awareness: Understanding the role of interpretation in assessment and in teacher learning. *Language Testing, 30*(3), 309-327. doi:10.1177/0265532213480128

Sheehan, S and Munro, S. (2017). *Assessment: attitudes, practices, needs*. British Council: London

Sheehan, S and Munro, S. (2019). *Classroom assessment: the development of teachers' cognitions*. British Council: London

Vogt, K and Tsagari, D. (2014). Assessment Literacy of Foreign Language Teachers: Findings of a European Study. *Language Assessment Quarterly, 11*(4), 374-402. doi:10.1080/15434303.2014.960046

Xu, Y and Brown, G. (2016). Teacher assessment literacy in practice: A reconceptualization. *Teaching and Teacher Education 58*. doi:10.1016/j.tate.2016.05.010

Yin, M. (2010). Understanding Classroom Language Assessment Through Thinking Research. *Language Assessment Quarterly, 7*(2), 175-194. doi:10.1080/15434300903447736

THÁI ĐỘ CỦA GIÁO VIÊN VỀ ĐÁNH GIÁ VÀ ẢNH HƯỞNG CỦA THÁI ĐỘ GIÁO VIÊN ĐỐI VỚI THỰC TIỄN ĐÁNH GIÁ TRÊN LỚP – KẾT QUẢ CỦA MỘT KHẢO SÁT TOÀN CẦU

Susan Sheehan

Khoa Giáo dục và Phát triển Chuyên môn
Đại học Huddersfield, Queensgate, Huddersfield

Tóm tắt: Bài viết này nghiên cứu thái độ của giáo viên đối với đánh giá. Thái độ của giáo viên đã được nghiên cứu thông qua "Khung nhận thức của giáo viên ngôn ngữ" của Borg (năm 2015). Nghiên cứu này áp dụng phương pháp nghiên cứu hỗn hợp sử dụng bảng câu hỏi và phỏng vấn. Một cuộc khảo sát trực tuyến đã được thực hiện với các giáo viên dạy tiếng Anh tại 57 quốc gia khác nhau trên toàn thế giới. Bảng câu hỏi được chia thành 3 phần. Phần thứ nhất bao gồm các câu hỏi liên quan đến kinh nghiệm đánh giá tại trường của những người tham gia nghiên cứu. Phần thứ hai tìm hiểu kinh nghiệm đào tạo đánh giá cả trong giảng dạy ban đầu và trong tất cả các khoá phát triển chuyên môn họ đã tham gia. Phần thứ ba và cũng là phần cuối cùng xem xét thực tiễn đánh giá của những người tham gia và quan điểm của họ về đánh giá. Ba phần này đều dựa trên lý thuyết "Khung nhận thức của giáo viên ngôn ngữ". Các bài phỏng vấn được thực hiện nhằm tìm hiểu lý do đằng sau các câu trả lời trong bảng câu hỏi. Quá trình phân tích dữ liệu cũng dựa trên khung lý thuyết. Kết quả nghiên cứu cho thấy kinh nghiệm giảng dạy trên lớp và các khóa phát triển chuyên môn có ảnh hưởng lớn nhất đến thái độ của giáo viên. Giáo viên bị ảnh hưởng bởi kinh nghiệm đánh giá của họ ở trường vì họ tránh lặp lại các phương pháp đánh giá đã được sử dụng để đánh giá họ.

Từ khóa: Đánh giá, nhận thức giáo viên, bảng câu hỏi

TERTIARY LANGUAGE PLANNING AND CAREER DEVELOPMENT: CHALLENGES TO VIETNAMESE STUDENTS

Nguyen Thi Thom Thom*[1], Bui Thi Ngoc Thuy[2], Nguyen Duc An[3]

1. VNU University of Languages and International Studies, Pham Van Dong, Cau Giay, Hanoi, Vietnam
2. Hanoi University of Science and Technology, Dai Co Viet, Hai Ba Trung, Hanoi, Vietnam
3. Taybac University, Quyet Tam, Sonla, Vietnam

Received 21 September 2019
Revised 08 May 2020; Accepted 23 July 2020

Abstract: Research in the field of English language education has shifted towards interrogating the real role of English in development across disciplines and countries. Vietnam English language policy, especially the National Foreign Languages Project (The NFL Project), have promoted English as a key success factor for Vietnamese students in the labour market as well as the well-being of the nation. This paper focuses on examining the notion of development which investigates the roles of English in Vietnamese tertiary students' employability. The researchers conducted 527 surveys with tertiary students in four university in the North Vietnam. Moreover, five students and three EFL tertiary instructors from each institution were selected to participate in a follow-up semi-structured interview (N = 32). The results indicated students' limited level of confidence in both their English skills as well as career-related skills performed in English. In other words, students showed their lack of sufficient English for employability purposes. Aligning the research results with the promising and well-intended agendas of ELP and The NFL Project illuminates a clear mismatch between English and employability. In reality, the relationship between English language education and career development is highly complex and contested, which may influence students' full development in the long run. Towards the end, the paper offers suggestions for improving pedagogies, policies and practices to promote English as one of vital employability skills for multifaceted personal and national developments.

Keywords: Employability, English, Notions of Development, Tertiary Education

1. Introduction

The field of English language education (ELE) has been increasingly gravitated towards uncovering the role of English in enhancing students' successful employment (British Council, 2014), especially in many Asian countries where the diversification of the economy is increasing (Erling, 2014; Coleman, 2011). Leaners are required to learn English for desirable jobs and other social and political capitals (Aslam, Kingdon, & Kumar, 2010; Grin, 2001; Ku & Zussman, 2010). English for career development has been increasingly scrutinized in many countries such as Afghanistan, Bangladesh, India, Iran, Nepal, Pakistan, and Sri Lanka (Coleman,

* Tel.: 84-912085632
Email: ThiThomThom.Nguyen@uon.edu.au; thomthomnguyenthi@yahoo.com

2011; Erling & Seargeant, 2013). Aligning with this research trend, there is a great need to uncover the relationship between English language policy planning and Vietnamese tertiary students' employability in light of the recent extensive ELE reforms in Vietnam.

The influential expansion of English created momentum for various Vietnamese language policy shifts. These signification transformations greatly emphasize English as a competitive skill for both personal and national advancement. English is mandated for all students across the education levels. According to Nguyen (forthcoming), English education has been a key role for university students to harness success and privilege in the world of work.

Using a part of data of a recent longitudinal mixed-methods study in four universities in different regions across North Vietnam, this paper attempts to uncover the influence of the current Vietnamese English language policies in facilitating tertiary students' employability. The first data set was collected from September to December 2016. As many as 527 full-time undergraduate students in 20 different majors such as Nursery Education, Agriculture, Electro-mechanics, Information Technology, Biology-Chemistry, and Tourism (See Table 1) were invited to conduct a survey. Most of these students (97.9 %) had seven-year experience of learning English before enrolling in their current university. As the requirement of their institutional curriculum, they were required to take 14-credit-hour English courses during their first and second year. They had to meet such credit number for graduation. The data was enriched by a set of qualitative components, that is, follow-up semi-structured interviews with three EFL tertiary instructors and five students from each of the four universities (N = 32). Each interview lasted from 15 to 20 minutes in the interviewee's suggested place and was in Vietnamese.

Table 1: General information on the research sites and participants

University	N	Majors of students surveyed
University 01	123	Sewing and Fashion Technology, Accounting, Economics, Information Technology
University 02	142	Physics, Mathematics, Mathematics-English Pedagogy, Information Technology, Primary Education, Nursery Education, Tourism, Vietnam Studies – Cultural Tourism
University 03	132	Agriculture, Economics, Electro-mechanics
University 04	130	Economics, Biology-Chemistry, Mathematics, Information Technology, Environment

Our preliminary data analysis addressed both EFL tertiary instructors' and students' perceptions of students' employability capacity. Students tended to show their low level of confidence in their English skills for employability as well as their ill-prepared command of career-related skills performed in English. Based on a review of the Vietnamese policy frame for tertiary ELE and employability and the notions of development, we discuss a mismatch between tertiary students' English language learning and employability. When current tertiary English programs fail to address students' needs for improving their English skills and soft skills required for their future careers such as debating and presenting in English, students can hardly take full advantage of their potentials, leading to questions of how they can build a secured foundation for their individual future and multifaceted developments, as well as how they can contribute to the goal of national development as targeted in the policy.

This paper is organised in two sections and shaped by three related arguments. We conduct a literature review of the current policy for tertiary ELE and employability, then discuss the application of the notions of development to Vietnamese language policies. Our further discussions address (1) misalignment between English language programs and employability, (2) misalignment between English language programs and career-related skills/soft skills in English, and (3) recommendations for facilitating students' employability, as presented in the sections that follow.

2. Tertiary English language education and employability

Since the time of Doi Moi (Renovation), significant political and economic shifts of the nation have called for various English language policy reforms in Vietnam. Doi Moi decentralises the state's power over private and local enterprises and foster multi-dimensional domestic and international collaborations to save the country from potential economic isolation after a ten-year period (1975-1985) under the socialist-oriented market economy and consequences of the US trade embargo. Doi Moi entails a less authoritarian, more multifaceted, multi-segmented model of governance (McCargo, 2004; Gainsborough, 2010) that has facilitated positive education and language reforms, including the rapidly growing demand to learn English in Vietnam (Bui & Nguyen, 2016; Bui, Nguyen & Nguyen, 2018). In 2008, the government launched The National Foreign Languages Project 2020 (hereafter referred to as The NFL Project) as a breakthrough in ELE and in national language reform by far. The NFL Project calls for a great demand for educating young Vietnamese citizens with English language proficiency in order to prepare for the strong development of both national and regional employability, as well as job competitions, especially in the context of the Association of Southeast Asian Nations' (ASEAN) economic integration (Government of Vietnam, 2008).

This policy offers great innovative strategies and action plans in ELE across all education levels and disciplines. To head for the English expansion and ELE reforms, the country has fostered a series of top-down transformational activities ranging from establishing national language testing centres, developing intensive English programs for both teachers and students, and providing intensive courses in English for specific majors such as engineering, tourism, business, finance, and technology (Plan 808/KG-BGDĐT, 2012). Regarding English language pedagogy, the national language policy reforms encourage student-centred approaches in the English curriculum, stimulating students' creativity, autonomy and engagement in the learning process. The NFL project aims at the newly arisen context of a competence-based approach in assessment, which requires university graduates to achieve a certified command of English, equivalently ranging from B1 – CEFR for English non-majors – to C1 for English majors (Bui, Nguyen & Nguyen, 2017). Provision of English language programs and standardised qualifications at institutional levels target at catering for a greater demand of both domestic and international labour forces. This policy has also resulted in innovative curriculum and English language teacher education, especially re-evaluating EFL teacher training and re-training programs and provision of professional development for both in-service and pre-service EFL teachers (Nguyen, forthcoming).

Albeit motivational and powerful, this policy seems ambitious in its objectives, thus placing a vague effect on the students' employability. After eight years of implementation, the NFL Project was reported not to successfully reach its initial short-term goals, with much less than 50% of the set tasks being fulfilled. Representatives of the Ministry of Education and Training admitted that most English language teachers were under-standardised. Teachers' poor instructional quality seemingly resulted in students' poor English competencies, which could hardly satisfy the purposes of employability (Luong, 2016). The goals of the NFL Project; accordingly, have been re-considered with an extended timeline up to the year 2025, with newly lesson-drawn objectives and updates aiming to boost English for employability for 100% of non-English major graduates, 90% of vocational students, and 60% of employees in public sectors (Hoang, 2016). These facts call for a stronger evaluation and greater emphasis of the connectedness between ELE education and employability, stated in an official adapted version of The NFL Project called Project 2025, issued by the Vietnamese government in late December 2017. Grounded on The NFL Project in terms of general goals, Project 2025 aims to address three major aspects as follows: (1) renovating foreign languages education in the national education system, continuing to implement new foreign language programs at all levels, and improving learners' foreign languages competence to meet the study and work demand; (2) strengthening the competitiveness of human resources in the integration period, contributing to the purpose of national development; and (3) establishing a nationwide foreign language foundation for general education in 2025 (Government of Vietnam, 2017).

While the intensive focus on English and ELE has emerged both in Vietnamese contexts and globally, a wide range of researchers in the field express their great concern for the significance of English for tertiary students' career development (Ferguson, 2013; Phillipson, 2012; Seargeant & Erling, 2013; Tollefson, 2013). Research into the impact of ELE indicates that the notion of English in development is contested and controversial. Thus, for an insight into the role of the current ELP to the students' development, we present the notion of development as a theoretical foundation for our study.

3. Applying the notion of development to Vietnamese English language policies

The socio-economic and political developments have greatly transformed the notion of development which is now to be considered as contested, inclusive, and multi-layered (Coleman, 2010). Aspects such as social and economic fulfillments, good governance, human rights, welfare, freedoms of social and economic engagement, employability, health care, and civil rights have become an epi-center in development (e.g., Coleman, 2011; Bruthiaux 2002). Development addresses the nation's ability to accommodate voices and choices and people's participation in social and economic planning and interventions (Seargeant & Earling, 2013).

Development and education have an intrinsic linkage as education promotes the knowledge economy for national development. Specifically, language proficiency and literacy programs set a solid foundation for advancing a skilled and interdisciplinary work force for development, which significantly fosters wide-ranging economic and social advancements, giving rise to high income, comprehensive social welfare, and human right protections (Coleman, 2010; Kirkpatrick & Sussex, 2012). Similarly, English is considered to possess a vital literacy tool for improving socio-economic and cultural capital (Seargeant & Earling, 2013).

Notions of development offer a crucial foundation for the interpretation of employability development in the current Vietnam English language policies. Development, as defined in this paper, offers a comprehensive insight into not only economic development but also a process to secure one's privileges and rights in various aspects of life including health, education, and social well-being. The concept of development is utilized to refer to implications as to whether the current English language policies prepare students with sufficient linguistic competencies for successful employment and personal advancements or whether such English language policies greatly bar them from employability capacities. The paper assumes that should tertiary students be equipped with good English language skills, they can be better prepared for their future employability and development. Looking from a nuanced perspective, students' English language proficiency can facilitate them to gradually gain an access to successful employment and social, educational, and economic negotiations. We further discuss this assumption in the sections of findings and discussion that follows.

4. Misalignment between English language programs and employability

A key finding of our research is that while students were all required to attend English courses at their institution, they did not benefit much from the institutional English language programs for employability skills. The majority of tertiary students (from over 70% to 80%) were not confident of their abilities to accomplish various activities in English such as listening to different topics (14.04% and 67.74% for "very unconfident" and "unconfident", respectively), collecting information (29.41% and 48.01%), answering phone inquiries (20.49% and 58.44%), and ability to understand human resource documents (22.96% and 54.65%) in English (See Appendix A). Students also lacked confidence in handling other activities required at their future workplace including understanding schedules, reading comprehension, reading information on the company websites, and creating documents, letters, reports, and invoices in English.

Justifications for these tertiary students' limited English for employability skills could include a range of factors from the institutional English programs to the students' learning related issuses. Both the EFL tertiary instructors and students interviewed shared that students could not communicate in English effectively or perform different tasks in English because communicative teaching approach is almost absent in their institutional curriculum. One student of economics major indicated a reality of his English:

I am not confident with my English for employability because our English program in Vietnam does not focus on educating students for communicative skills but on grammar and theories. Moreover, I did not learn English systematically when I was young, so I am afraid of re-learning it. *(Em không tự tin ... vì chương trình Tiếng Anh ở [nước] mình không tập trung vào các kỹ năng giao tiếp mà chỉ toàn ngữ pháp và lý thuyết. Em không được học Tiếng Anh bài bản từ bé nên em ngại học lại.)*

A EFL tertiary instructor in Agriculture major commented that even when the communicative approach was mentioned in their English program, there were no details of specific communicative tasks and assessments. What the teachers and students could mostly do in the classroom was focusing on the reading tasks, completing the vocabulary and grammar exercises. The instructor further explained that this English program could be most suitable to students of low levels, and other factors such as the time allotment in the syllabus and big class size did not allow for promoting students' communicative skills.

Many other students faced the same challenge in communicating in English because they lacked both confidence and motivation, and they found it hard to respond naturally. Consequently, students did not own a sense of preparedness for any job environments which require English. A student majoring in Math explained his pessimistic perspective: "I am not confident with my English for job related purposes in the future because my English skills are not enough for communication"/ *Em chẳng thấy tự tin chút nào với vốn Tiếng Anh để làm việc sau này vì đến tiếng Anh cơ bản em cũng chẳng đủ cho giao tiếp thông thường.)* He stated that:

"I have learnt English for nearly ten years, but not continuously. Sometimes, I just revised some English grammar for tests and examinations. When I attend my English class, I mostly sit silently, try to copy words or anything the teacher writes on the board. I cannot remember all the words and I hardly speak out any sentences in English. I can only do some word-by-word translation from English to Vietnamese. I'm ashamed of my English and my English language learning but I don't like learning English. At the moment, I feel it is not useful for my teaching Maths in the future." *(Em học Tiếng Anh gần 10 năm nay, nhưng mà không liên tục. Đôi khi em chỉ học ôn một chút ngữ pháp để chuẩn bị cho thi cử kiểm tra. Khi đi học giờ Tiếng Anh, em chủ yếu ngồi yên, cố chép từ và những gì thầy cô viết trên bảng vào vở cho có. Em chẳng nhớ nổi các từ đó và em chẳng nói nổi câu Tiếng Anh nào ra hồn. Em chỉ dịch từng từ một từ Tiếng Anh sang Tiếng Việt. Em xấu hổ với vốn Tiếng Anh và việc học Tiếng Anh của mình, nhưng em chẳng thích học Tiếng Anh. Giờ em thấy nó chưa hữu ích cho nghề nghiệp dạy Toán của em sau này.)*

The students' responses greatly coincided with the EFL tertiary instructors' perspectives that their students' English competence was not sufficient enough for future employability. These tertiary instructors expressed their worry about students' capability to function at an English-based workplace. They reflected that most students admitted their poor command of English, and "only a small number of students have been prepared with enough English knowledge and skills for English-related jobs"/ *chỉ số ít sinh viên đủ kiến thức và kỹ năng Tiếng Anh chuẩn bị cho các công việc có Tiếng Anh sau này).* As a matter of fact, almost all students are supposed to go back to their university or attend other language courses to "relearn English" to meet their job requirements. An English instructor in IT major mentioned a fact that at the beginning of any English course, she often advised her students to make effort in learning English by both following the institutional program and attending extra classes. She emphasized "The institutional program cannot fully facilitate your English improvement, you need to invest more to cope with job requirements in the near future." *(Chương trình học ở trường không đủ giúp các em tiến bộ với Tiếng Anh đâu. Các em cần đầu tư hơn nữa để bắt kịp với yêu cầu nghề nghiệp trong tương lai không xa.)* Another fact was that her students often 'rushed' to her for advice when they nearly completed the English program at the university or before graduation. These students tended to suffer from the pressure of having an urgent English proficiency certificate for their job application.

It seems that the English programs played an invisible role in helping both the EFL tertiary instructors and students feel confident about students' English for employability. Such uncertainty was rather complex on multiple issues including program development, students' unsystematic English language education, lack of environment to practise communicative English as well as students' low motivation and poor awareness to learn the subject.

5. Misalignment between English language programs and career-related skills

Similar to disconnect between English language programs and employability, institutional English language programs seemed to mismatch with the set task of preparing for students' career-related skills performed in English. To be specific, students' level of certainty about performing career skills in English was rather low. Student participants indicated they were largely incapable of debating (25.81% and 54.08%), solving problems (22.2% and 55.22%), and working with people worldwide (26.57% and 50.47%). Apart from these activities, a majority of students were unable to perform many other career-related skills including interpersonal skills, using technology, leadership, and searching for scholarships in English (See Appendix A)

Both EFL tertiary instructors and students indicated that various challenges hindered them from achieving the aim of improving students' above career-related skills in English. While interviewed instructors reflected that students "are idle, inactive, lazy, demotivated" and "just smile, do not want to talk", many students realised their problems with the lack of vocabulary to express their ideas and their fear of making mistakes. Consequently, students' performance of skills such as problem solving and group work performed in English only ranged from insufficient to average level. A nurse – major instructor's response stated:

Generally speaking, the students seem inactive and demotivated. Even when they know that English is important for employability, they do not seem to worry much. They do not try hard enough to improve their skills. Many do not complete the homework, prepare for the lesson or actively engaged in in-class activities. They think the English program at the university is just a subject. They wish to learn less yet achieve higher scores. *(Sinh viên nói chung là thụ động và thiếu động lực. Kể cả khi biết tiếng Anh rất quan trọng cho nghề nghiệp sau này, dường như các em cũng không tỏ ra lo lắng lắm. Các em chẳng cố gắng mấy để cải thiện kỹ năng. Nhiều em không làm bài về nhà, không chuẩn bị bài mới hay không tích cực tham gia các hoạt động trên lớp. Các em nghĩ chương trình học Tiếng Anh ở trường [đại học] chỉ là một môn học)*.

One EFL tertiary instructor in agriculture major reflected that the students could not work in groups effectively so sometimes she and her colleagues had to require them to prepare beforehand the topics in both Vietnamese and English. She observed that the students were not willing to talk or discuss. With problem solving skill, students did not even try to brainstorm or think because they got familiar with lecturers providing options for them. Therefore, "they can't think about solutions and even if they have some solutions, they are unable to express them in English". (*Vì vậy, sinh viên không thể nghĩ ra giải pháp và kể cả khi có giải pháp, họ cũng chẳng biết diễn đạt thế nào bằng Tiếng Anh).*

Addressing the students' presentation skills, this EFL tertiary instructor further commented that "students do not want to talk because of their limited vocabulary range. They speak slowly and have many repetitive phrases or sentences." (*Sinh viên không muốn tham gia vì vốn từ hạn hẹp. Các em nói rất chậm và thường hay lặp lại các cụm từ và câu).* Besides, not all students could have opportunities to practise their presentation skills because of limited time and lack of guidelines in the course syllabus.

Students also expressed their deficiency in communicating in English with people within and outside of their future work setting. They acknowledged that both their English and communicative skills were inadequate, so they might hesitate or avoid talking to others. When required to speak, they tended to feel nervous and even resorted to broken language or the type of minimal English that they had picked up after years of learning. Sharing the same perspective, almost all the instructors cast doubt on the students' ability to communicate in English in their work setting. An EFL tertiary instructor in business management major voiced his opinion:

Although students may have learnt English for many years, their English does not always improve when they reach tertiary levels. They become less open to speak. They are shy and cannot express themselves clearly and precisely. I think students can only function some simple conversations in English. Only a small number of students can work effectively with people in the region and worldwide in English. *(Dù sinh viên có thể học tiếng Anh nhiều năm nhưng trình Tiếng Anh cuả các em không phải lúc nào cũng tiến bộ khi lên đại học. Các em ít 'open' hơn khi phải nói, ngại và không biết cách diễn đạt ý của mình rõ ràng và chuẩn xác. Mình cho rằng sinh viên chỉ nói được vài dạng hội thoại cơ bản bằng Tiếng Anh. Chỉ có số ít sinh viên có thể làm việc hiệu quả bằng Tiếng Anh với người trong khu vực và quốc tế).*

Students' lack of English competencies may lead to their inability to perform soft skill activities such as debating, presentation, working in groups, and showing their role as a leader in English effectively. The justifications can be students' English learning and EFL teaching pedagogies embedded in their English programs. Thus, it may also be implied from this study that ways of doing education at the tertiary level should be addressed. Details of such implications will be presented in the final part of this paper.

6. Recommendations for facilitating students' employability

Several suggestions from both the students' and lecturers' perspectives were proposed to enhance students' English as an attribute critical to employers. On that ground, our study further recommended a great emphasis on appropriate pedagogy, innovative language programs, and the provision of skill-based courses for employability. This can better suit students' potentials and needs and capture their attention to learn English successfully. Discussing how to improve institutional language program, a student voiced that: 'We want to learn English systematically from basic to advanced levels, following a clear guideline. We are ready to start from scratch." *(Chúng em muốn học Tiếng Anh hệ thống từ cơ bản đến nâng cao, the chỉ dẫn rõ ràng. Chúng em sẵn sàng học [lại] từ đầu.)* Another student majoring in Electro-Mechanics added that they need practise and drill their communicative skills in different contexts. Mentioning the current university programs, both these student participants agreed that they would like to be provided with more practical and interesting lessons and that teachers should apply "updated" teaching methods. Of a particular note, offering appropriate teaching methods with the performance-based approach to suit students' needs and ages was highly emphasized. Further, such methods may equip students with soft skills which enable them to function more successfully and encourage their active learning performance.

Along with the improvement in quality of instruction, teaching communicative skills, namely listening and speaking skills should be oriented and spent more time to redress the balance of all four language skills because the time allocated for the current English language course at secondary levels was mainly for reading and some basic writing. There should be more extracurricular activities such as

English clubs and Guest speakers with the participation of invited employers which can provide students a wide range of opportunities for both their English and English-related skill enhancement. More importantly, inclusion of diverse soft skill courses and job searching workshops in current institutional English programs would better prepare students for their future employability. Many of the student interviewees expressed their need to be provided with more specific and hands-on skill courses. One student emphasised that they were in dire need of practising soft skills such as communication, debating and technology.

In addition, according to the interviewed EFL tertiary instructors and students, a program development division in each university and/or an employability counselling unit should be established to focus and incorporate companies' demands with the learning outcomes. As a result, the lecturers' burden of time constraints can be alleviated because this division can provide them with ‘‘employment scenarios’’ and information on the labour market to be included in their teaching. To the suggestion of a lecturer of automobile major, English for job interviews should be added to her institutional curriculum and assessment tools should be improved to better the outcome-based assessment. To illustrate, while the four language skills are integrated into teaching and learning process to enable students to be more confident and successful in real-life situations, the formative assessment focusing on students' learning process should be used as the main tool instead of the summative which is widely used. What is more, students' language communicative competence should be officially tested, instead of the current lexico-grammar oriented exams. To a lecturer of Economics, both teachers and test designers need to redress the balance of all four skills with the association with checklists of "I can" descriptors.

Collected data figured out that students' attitudes and reactions to the importance of English language in their future jobs had resulted in their poor attainment and almost illiteracy in English after over seven years of academic exposure to English at secondary levels. The most common explanations provided were that English seemed to be an impossible subject for its difficulty, there were no improvements in their language competence, and their motivation of English language learning is too low as they lived in remote areas with no or very little contact with English speakers and they learned English for doing nothing. As such, both the importance of learning English and promoting the habit of learning English among students should be highlighted in a persistent and systematic manner so that students' employability skills can be ultimately enhanced.

The quality improvement of English language teaching and learning would be extremely challenging and complex in the context of this study since it requires the involvement of managers, policy makers and educators of all levels. Such recommendations further imply that many aspects including teaching methods/quality of instruction, program development, curriculum design, students' needs, and employers' requirements need to be taken into serious consideration to issue appropriate and feasible policies, decisions and requirements, hence, aiming at improving English for students' career development.

7. Discussion and conclusion

The study aims to address the under-researched students' English language skills for employability. The results show the increasingly limited role of English in fostering students' successful employment because the English language policies largely fail to equip students with good English competency. Students seem to be unable to use their English for as one of the tools for employability demands set by employers, their society, and beyond. English is obviously, unable to function as a linguistic capital, sanctioning students with power and dispositions in their social world.

The results further suggest that the English programs seem to be loosely structured and administered. Such policies could de-facto stealing students' time while de-motivating their passion towards having solid English language skills. In other words, such language programs in this context may close students' multiple opportunities later in life when teaching and learning activities still seems to be incomprehensive, contested, and controversial.

The study also aims at providing some rigorous recommendations for enhancing students' English and employability capacities.

First, a systematic English education from primary level onwards is recommended. This is because many students in this study possess a poor English proficiency although they have learned English for such a long time. They need to have a solid and systematic English program from lower levels, which could set a good foundation for their English in tertiary education. University students face a doubled burden because they have to learn both their major and English. However, they seem to struggle with English rather than other subjects. This is the reason why students were not confident in performing a job using English.

Second, students need to be guided to a mindset of life-long learning when it comes to learning a language. Learning language takes a tremendous amount of time and effort and they need to be mentored to learn in a strategic and persistent manner. We suggest lecturers to teach students learning how to learn, self-studying while mobilizing the advancement of technology in learning English (Bui, 2016). It is vital to take advantage of abundant online materials for their English and soft skill improvements (Darling-Hammond, 2010).

Finally, we need a collaborative and supportive effort between universities and employers. Career development divisions work with employers to tailor English programs which address employers' demands regarding skills and knowledge needed. Such divisions foster a ground for policy makers, lecturers, and administrators to structure and develop diverse courses to boost students' successful employment (Ku & Zussan, 2010; Lee, 2012). Career-orientation workshops and internship programs should be promoted to better prepare for labour market requirements while enriching students' hands-on experiences and relationships with employers (Pooja, 2013). We highly agree with Bui, Nguyen, and Nguyen (2018; p.65) to argue that "English as employability skill for development should be developed in a systematic, collaborative, diverse, and conscientious approach to meet ever-demanding requirements of both domestic and international employers." We need tremendous effort from multiple stakeholders: the students themselves, policy makers, the tertiary governance, and employers to develop a sustainable workforce. Successful employment contingents on comprehensively and rigorously structured language policy planning and practices.

References

Aslam, M., De, A., Kingdon, G. & Kumar, R. (2010). *Economic Returns to Schooling and Skills - An Analysis of Pakistan.* Cambridge, UK: University of Cambridge.

Bruthiaux, P. (2002). Hold your courses: Language education, language choice, and economic development. *TESOL Quarterly*, *36*(3), pp. 275-296.

Bui, T. T. N., & Nguyen, T. M. H. (2016). Standardizing English for educational and socio-economic betterment: A critical analysis for English language policy reforms in Vietnam. In R. Kirkpatrick

(Ed.), *English language education policy in Asia* (pp. 363-388). Switzerland: Springer International Publishing.

Bui, T. N. T., Nguyen, T. T. T., & Nguyen, D. A. (2017). English as a Key Employability Capacity: Perspectives from Vietnamese Students and Lecturers. *Language Education in Asia, 8*(2), pp. 144-156.

Bui, T. N. T., Nguyen, T. T. T., & Nguyen, D. A. (2018). Vietnamese higher education language planning and university students' career development. In J. Albright (Ed.), *English Tertiary Education in Vietnam* (pp.54-67). London: Routledge.

British Council. (2014). *Annual report 2013-14*. London: British Council

Coleman, H. (2010). *The English Language in Development*. London: British Council.

Coleman, H. (Ed.). (2011). *Developing countries and the English language: Rhetoric, risks roles and recommendations*. London: British Council.

Darling-Hammond, L. (2009). *The flat world and education: How America's commitment to equity will determine our nation's future*. New York: Teachers College Press.

Darling-Hammond, L. (2010). Teacher Education and the American Future, *Journal of Teacher Education*, 61(*35*). Doi: 10.1177/0022487109348024

Erling, J. E., & Seargeant, P. (2013). *English and development: Policy, pedagogy, and globalization*. Bristol, UK: Multilingual Matters.

Erling, E. J. (2014). *The roles of English in skills development in South Asia: Policies, interventions and existing evidence*. London: British Council.

Ferguson, G. (2013). English, development and education: Charting the tensions. In E. Erling & P. Seargeant (Eds.), *English and development: Policy, pedagogy, and globalization* (pp. 21-44). Toronto, Canada: Multilingual Matters.

Gainsborough, J. (2010). *Vietnam: Rethinking the state*. London and New York: Zed Books and Silkworm Books.

Government of Vietnam. (2008). *Decision 1400/QD-Ttg on the approval of the project "Teaching and Learning Foreign Languages in the National Education System, Period 2008–2020"*.

Government of Vietnam. (2017). *Decision 2080/QD-Ttg on the refresh of the project "Teaching and Learning Foreign Languages in the National Education System, Period 2017–2025"*.

Grenfell, M. (2012). Bourdieu, language and education. In M. Grenfell, D. Bloome, C. Hardy, K. Paul, K. Rowsell, & B. Street (Eds.), L*anguage, ethnography and education: Bridging new literacies and Bourdieu* (pp. 50-70). New York: Routledge.

Grin, F. (2001). English as economic value: facts and fallacies. *World Englishes*, *20*(1), pp. 65-78.

Hamid, O. (2015). *English and employability, mobility and development*. Retrieved from www.britishcouncil.org.bd/sites/default/files/bc_think_piece_obaid_hamid_08_10_2015.pdf

Hanks, W. F. (2005). Pierre Bourdieu and the practices of language. *Annual Review of Anthropology*, *34*(1), 67-83. doi:10.1146/annurev.anthro.33.070203.143907

Hoang, P. (2016). *Đề án ngoại ngữ gần 9400 tỷ sau 8 năm làm được những gì*. Retrieved from http://vnexpress.net/tin-tuc/giao-duc/de-an-ngoai-ngu-gan-9-400-ty-sau-8-nam-lam-duoc-nhung-gi-3470788.html

Kirkpatrick, A., & Sussex, R. (2012). *English as an international language in Asia: Implications for language education*. New York: Springer.

Kramsch, C. (2008). Pierre Bourdieu: A biographical memoir. In J. Albright & A. Luke (Eds.), *Pierre Bourdieu and literacy education* (pp. 33-49). New York: Routledge.

Ku, H., & Zussman, A. (2010). Lingua franca: The role of English in international trade. *Journal of Economic Behavior and Organization*, *75*(2), pp. 250-260.

Lee, C. G. (2012). English language and economic growth: Cross-country empirical evidence. *Journal of Economic and Social Studies*, *2*(1), pp. 5-20.

Luke, A. (2008). Using Bourdieu to make policy: Mobilizing community capital and literacy. In J. Albright & A. Luke (Eds.), *Pierre Bourdieu and literacy education* (pp. 347-361). Malwah, NJ: Lawrence Erlbaum Associates.

Luong, H. N. (2016). *Đề án ngoại ngữ quốc gia 2020 thất bại vì đâu.* Retrieved from http://baoquocte.vn/de-an-ngoai-ngu-quoc-gia-2020-that-bai-vi-dau-39673.html

McCargo, D. (2004). Introduction. In D. McCargo (Ed.), *Rethinking Vietnam* (pp. 1-12). London: Routledge.

Nguyen, T. T. T. (forthcoming). *Professional development: Vietnamese EFL teacher educators' perceptions and experiences*. (Doctoral Dissertation), The University of Newcastle, Australia.

Phillipsons, R. (2012). Linguistic imperialism alive and kicking. *The Guardian*. Retrieved from www.guardian.co.uk/education/2012/mar/13/linguistic-imperialism-english-language-teaching

Plan 808/KH_BGDDT. (2012). *Kế hoạch triển khai đề án ngoại ngữ 2020 trong các cơ sở giáo dục đại học giai đoạn 2012–2020.* Retrieved from www.moet.gov.vn/?page=6.21&script=%20congvan&namxly=-1

Pooja, B. (2013). English for employability-A challenge for ELT faculty. *Research Journal of English Language and Literature*, *1*(3), pp. 350-353.

Seargeant, P., & Erling, E. J. (2011). The discourse of 'English as a language for international development': Policy assumptions and practical challenges. In H. Coleman (Ed.), *Dreams and* *realities: Developing countries and the English language* (pp. 248-267). London: British Council.

Seargeant, P., & Erling, E. J. (2013). Introduction: English and development. In E. Erling & P. Seargeant (Eds.), *English and development: Policy, pedagogy, and globalization* (pp. 22-45). Toronto, Canada: Multilingual Matters.

Tollefson, J. W. (2013). *Language policies in education: Critical issues* (2nd ed.). New York: Routledge.

ĐỊNH HƯỚNG NGÔN NGỮ VÀ PHÁT TRIỂN NGHỀ NGHIỆP: THÁCH THỨC VỚI SINH VIÊN VIỆT NAM

Nguyễn Thị Thơm Thơm[1], Bùi Thị Ngọc Thủy[2], Nguyễn Đức Ân[3]

1. Đại học Ngoại ngữ, Đại học Quốc gia Hà Nội, Phạm Văn Đồng, Cầu Giấy, Hà Nội, Việt Nam
2. Đại học Bách khoa Hà Nội, Đại Cồ Việt, Hai Bà Trưng, Hà Nội, Việt Nam
3. Đại học Tây Bắc, Quyết Tâm, Sơn La, Việt Nam

Tóm tắt: Xu hướng nghiên cứu trong lĩnh vực giảng dạy tiếng Anh những năm gần đây tập trung nhiều vào vai trò thực tế của tiếng Anh trong sự phát triển của các ngành nghề và các quốc gia. Chính sách ngôn ngữ về tiếng Anh ở Việt Nam, đặc biệt là Đề án Ngoại ngữ Quốc gia coi tiếng Anh như là chìa khóa thành công của sinh viên Việt Nam trong thị trường lao động cũng như sự phát triển bền vững của đất nước. Nghiên cứu này nhìn nhận lại các quan điểm về phát triển trong đó chú trọng đến vai trò của tiếng Anh trong việc phát triển các kĩ năng nghề nghiệp của sinh viên đại học Việt Nam. Khảo sát được thực hiện với 527 sinh viên hiện đang theo học tại 4 trường đại học ở khu vực phía Bắc. Sau đó, 5 sinh viên và 3 giảng viên tiếng Anh ở mỗi trường được lựa chọn để tham gia phỏng vấn (N = 32). Kết quả thu được cho thấy sinh viên không tự tin vào kĩ năng tiếng Anh cũng như các kĩ năng nghề nghiệp khác cần sử dụng tiếng Anh. Nói cách khác, sinh viên thiếu kĩ năng tiếng Anh phục vụ cho mục đích công việc. Đối chiếu kết quả nghiên cứu với lộ trình và mục tiêu của Đề án Ngoại ngữ Quốc gia có thể thấy rõ sự khập khiễng giữa năng lực tiếng Anh và các kĩ năng nghề nghiệp. Thực tế, mối quan hệ giữa giảng dạy tiếng Anh và sự phát triển nghề nghiệp khá phức tạp và gây nhiều tranh cãi, điều này về lâu dài ít nhiều ảnh hưởng tới sự phát triển toàn diện của người học. Từ những vấn đề quan trọng được chỉ ra từ kết quả nghiên cứu, nhóm tác giả đề xuất một số giải pháp liên quan đến việc phát triển các chính sách và cách thức thực hiện nhằm nâng cao vai trò và hiệu quả của tiếng Anh như là kĩ năng nghề nghiệp cần thiết đối với sự phát triển toàn diện của mỗi cá nhân và đất nước.

Từ khóa: kĩ năng làm việc, tiếng Anh, quan điểm về phát triển, giáo dục đại học

Appendix A - Students' English skills and English as an employability skill

	1 – Very unconfident		2 – Unconfident		3 – Confident		4 – Very confident		5 – Not sure	
	n	%	N	%	n	%	n	%	n	%
I English skills										
1 Ability to collect information	155	29.41	253	48.01	72	13.66	8	1.52	39	7.40
2 Ability to listen to different topics	74	14.04	357	67.74	51	9.68	7	1.33	38	7.21
3 Ability to communicate	108	20.49	281	53.32	60	11.39	26	4.93	52	9.87
4 Ability to answer phone inquiries	108	20.49	308	58.44	59	11.20	10	1.90	42	7.97
5 Ability to do presentations	115	21.82	253	48.01	80	15.18	22	4.17	57	10.82
6 Ability to introduce yourself	62	11.76	229	43.45	161	30.55	22	4.17	53	10.06
7 Ability to answer interview questions	113	21.44	285	54.08	59	11.20	19	3.61	51	9.68
8 Ability to comprehend readings	86	16.32	236	44.78	127	24.10	18	3.42	60	11.39
9 Ability to understand timetables and job-related agendas	96	18.22	264	50.09	86	16.32	13	2.47	68	12.90
10 Ability to read information regarding companies online	86	16.32	286	54.27	94	17.84	17	3.23	44	8.35
11 Ability to understand human resource policies	121	22.96	288	54.65	47	8.92	12	2.28	59	11.20
12 Ability to create documents, for example, letters, instructions, annoucements, forms, and email	112	21.25	271	51.42	66	12.52	10	1.90	68	12.90
II English as an employability skill										
13 Ability to work with colleagues in groups	105	19.92	281	53.32	71	13.47	13	2.47	57	10.82
14 Ability to solve problems	117	22.20	291	55.22	55	10.44	9	1.71	55	10.44
15 Ability to share, collaborate, support, and cooperate	96	18.22	295	55.98	69	13.09	13	2.47	54	10.25
16 Ability to use information technology	106	20.11	272	51.61	82	15.56	17	3.23	50	9.49
17 Ability to do business in English	85	16.13	272	51.61	93	17.65	16	3.04	61	11.57
18 Ability to function as a leader	112	21.25	268	50.85	72	13.66	14	2.66	61	11.57
19 Ability to debate	136	25.81	285	54.08	47	8.92	10	1.90	49	9.30
20 Ability to find good jobs	138	26.19	263	49.91	59	11.20	8	1.52	59	11.20
21 Ability to work with people worldwide	140	26.57	266	50.47	54	10.25	12	2.28	55	10.44
22 Ability to search for scholarships to study abroad to develop your professionalism	128	24.29	240	45.54	40	7.59	7	1.33	112	21.25

VNU Journal of Foreign Studies, Vol.36, No.4 (2020) 153-169

WASHBACK TO LANGUAGE TEACHERS: A REVIEW OF MODELS AND EMPIRICAL RESEARCH IN AND BEYOND VIETNAM

Dinh Minh Thu*

Haiphong University
171 Phan Dang Luu, Kien An, Hai Phong, Vietnam

Received 15 October 2019
Revised 23 April 2020; Accepted 24 July 2020

Abstract: Washback, i.e., test effects on teaching and learning, has been emerging as an attractive research topic in language training and assessment for over the past 20 years for its significant implications of test validation and fairness for both policy-makers and practitioners. Presently, it deserves more Vietnamese researchers' interest in the context of the enactment of the National Foreign Language Project 2020 (extended to 2025), which puts language assessment as a key innovation requirement. Washback operates either positively or negatively; i.e. promoting or inhibiting learning. Teachers are considered the precursor in the washback mechanism. There is only one washback model on the washback effects on teachers, which is proposed by Shih (2009). This paper aims to critically browse other washback models besides Shin's (2009) to generate a washback framework on teachers' perceptions and practices. Previous empirical washback research on teachers in and beyond Vietnam is, then, investigated in alignment with the aspects illustrated in the framework to point out achievements and gaps in the field. A qualitative approach of document analysis of over forty studies of differing types, i.e. books, dissertations and articles, has been adopted to reach the research aim. The discussion is divided into two major parts, including the washback models pertaining to teachers to scaffold a model for teachers' perceptions and practices, and the results in empirical research in terms of the aspects mentioned in the model. Findings show that washback on teachers' perceptions ranges from perceptions of the test itself, students' language ability, teaching contents and methodology to teachers' professional development. Plus, washback on teachers' practices concerns their selections of teaching contents and methodology in class as well as their involvement in professional development. The element of professional development can be considered a new light in the reviewed washback model. This has a significant meaning by raising teachers' awareness of developing themselves professionally. The current paper expects to contribute to elaborating the scenario of washback research for interested researchers, practitioners and policymakers not only in but beyond the context of Vietnam.

Keywords: washback, washback models, language test, teacher perceptions, teacher practices

1. Introduction

[1]Washback, i.e., test effects on teaching and learning, has been attracting numerous researchers like in the world, including Vietnam (Alderson an Banerjee, 2001; Bui, 2016; Bui, 2018; Cheng & Curtis, 2012; Nguyen, 2017; Hsieh, 2017; Tayeb, Abd Aziz & Ismail, 2018; Wall & Horák, 2006; Wenyuan, 2017). According to Cheng, Sun, and Ma (2015, p. 440), the popularity of washback was justified by its effect on test fairness and validation. It is undeniable that teachers are the precursor in the process of

* Tel.: 84-912362656
Mail: thudm@dhhp.edu.vn

teaching and training. This argument raises the need of studies on washback on teachers, who can create positive washback in class to promote learning. Documentation has recorded washback models proposed by Alderson and Wall (1996), Bailey (1996), Hughes (2003), Green (2007) and Shih (2009). Nonetheless, there leaves a gap of a single washback research review which updates the washback theories and empirical findings from the teacher aspect. The current study aims to fill into that gap by answering two research questions as follows:

1. What is the shape of the updated test washback model on English language teachers' perceptions and practices?
2. How have the aspects in the updated model been studied?

The research expects to provide a new look into the washback reseach area for English language assessment not only in Vietnam but beyond the country.

2. Methodology

The qualitative approach is applied to this review via a document analysis of the previous research on washback theories and practices. The literature was analysed and evaluated critically in accordance with the research questions. The research started with the definitions of washback, teachers' perceptions, and teachers' practices. The major research part embraced the critical revision of recognized washback conceptual frameworks, from which a new conceptual framework for washback to EFL teachers' perceptions and practices was built. Plus, it reviewed the empirical findings on the bases elaborated in the fresh framework.

3. Theoretical background

3.1. Washback concepts

Washback (*backwask)* has been largely defined in applied linguistics. The most general concept of washback can be "the effect of testing on teaching and learning" (Hughes, 2003, p.1). More specific concepts can identify the individuals involving in the washback mechanism or/and the context washback occurs in. Washback refers to "the impact of external language tests to affect and drive foreign language learning in the school context" (Shohamy, 1993, p. 153); "the direct impact of testing on individuals" (Bachman & Palmer, 1996, p. 30); the force for "teachers and learners to do things they would not necessarily otherwise do because of the test" (Alderson & Wall, 1993, p.1); or "a part of the impact a test may have on learners and teachers, on educational systems in general, and on society at large" (Hughes, 2003, p. 53). These definitions point out teachers, students and other stake-holders like authorities and parents who are affected by the test powers. Also shown from these concepts, washback can operate either "in the school context" (Shohamy, 1993, p.153) or even in the society (Hughes, 2003). In addition, Pierce (1992, p. 687) contributed to washback definitions by stating that it is "the impact a test has on classroom pedagogy, curriculum pedagogy, curriculum development and educational policy". His definition is interested in teachers and policy makers rather than learners through the words of "pedagogy" and "policy". Another interesting point of view on washback was Pearson's (1988, p. 7), cited in Cheng et al. (2004): "Public examinations influence the attitudes, behaviours, and motivation of teachers, learners, and parents, and because the examinations often come at the end of a course, this influence is seen as working in a backward direction, hence the term, washback." Pearson' (1988) point of view comprises both the cognitive features like attitudes and motivation and the practice or behaviour. This research concerns washback effects on teachers in the school context since "teachers are "the 'front-line' conduits for the washback processes related to instructions" (Bailey, 1996, p.17). The above analysis yields a clear shape of washback which means the test influence on teachers' cognitive mechanism and actions to reach the educational goals. This research conceptualizes washback as the classroom impact of tests on teachers' perceptions and practices toward teaching and learning.

3.2. Teachers' perception

Teachers' perceptions, one of the two focal points of the current study, have been mentioned in Alderson & Wall's (1993), Hughes' (2003), Green's (2007) and Shih's (2009) washback theory. However, little effort has been made to define in teachers' perceptions in relevance to washback effects. In empirical research on teachers' perceptions, the words of "perceptions" and "beliefs" are used interchangeably (Wang, 2010; Onaiba, 2013; Mahmoudi, 2013; Antineskul & Sheveleva, 2015; Cheng, 1999; Hsu, 2009; Liauh, 2011; Salehi et al., 2012; Cheng, 2004) without much effort in defining perceptions but beliefs.

With regards Cambridge Dictionary, *perception* is defined as "a belief or an opinion" or "an understanding". Instead of providing a thorough insight into perceptions, cognition researchers have widely discussed the term beliefs (Pajares, 1992; Borg, 2003; Zeng, 2015). There is inconsistency in defining teachers' beliefs. While Green (2012) and Richardson (1996) cited as Le (2011) distinguish beliefs from attitudes and knowledge, Borgs (2003) and Pajares (1992) consider beliefs knowledge, perceptions and attitudes. Then, perceptions can be understood through the definitions of beliefs. Rokeach (1969) as cited in Le (2011) sets beliefs as an "integrated cognitive system" or "any simple proposition . . . inferred from what a person says or does, capable of being preceded by the phrase 'I believe that …" Pajares (1992, p. 316) defines beliefs as an "individual's judgment of the truth or falsity of a proposition, a judgment that can only be inferred from a collective understanding of what human beings say, intend, and do". Richardson (1996, p. 102) names beliefs "a subset of a group of constructs that name, define, and describe the structure and content of mental states that are thought to drive a person's actions". Perceptions belong to these constructs. Borg (2003, 2006) states teachers' beliefs are the cognitive and systemic nature of beliefs: what teachers think, know or believe.

In washback research, teachers' perceptions are grounded on the label "attitudes", "feelings" (Mahmoudi, 2013; Tsagari, 2011), "beliefs" (Mahmoudi, 2003, Wang, 2010), "understanding" (Cheng, 2004; Hsu, 2009). Antineskul & Sheveleva (2015), reflected the research on teachers' perceptions with the words "attitude", "think", "like", and "know" repeated many times (p. 8 -12). Onaiba (2013, p. 56) accredits perception washback to feelings, beliefs, attitudes toward the test. Only Mahmoudi (2013) mentions perceptions and attitudes separately from the title of his research, and only Green (2013) talks about beliefs, not perceptions. Green (2013, p. 46, 47) raises specific questions on teachers' beliefs about teaching and about testing. Regarding teaching, they are teachers' beliefs of effective teaching strategies and their compatibility with the test demands, of test preparation challenges and of "local precedents" for that preparation.

From the above review, teachers' perceptions of teaching under the influence of the test denote how teachers feel, think about, believe and understand the test and their classroom teaching practices.

3.3. Teachers' practices

Almost all previous empirical washback studies have excluded the review of teachers' practices, but perceptions. Barnes (2017) seems to be the single washback researcher who discusses the relationship between communicative language teaching (CLT) and high-stakes language testing prior to the methodology part and other subsequent parts. Hsu (2009) provided "teachers' behaviors" as "what teachers do in the classroom" (p.88), and he studied teachers' medium of instruction, teacher talk, teaching activities, teaching materials and lesson planning. The deficiency in definitions of teachers' practices in washback research may imply the researchers find tests in teaching rather than teaching in tests. It can be argued that when relevant teaching theories are discussed in a washback study on teachers' perceptions of teaching and their actions, from which teaching aspects come into lights to facilitate the evaluation of teaching effectiveness in the introduction of a new test.

Concerning teachers' effectiveness, Danielson (1996) presents a teaching framework of four domains, including Planning and Preparation, Classroom Environment, Instruction and Professional Development. The three first domains concern teachers' direct actions in class, while the fourth and last domain enhances the quality of direct actions. The planning and preparation section requires teacher knowledge of content, methodology, students, resources and assessment. The second domain pertains to teachers' ability to creating and managing a class which fosters learning. The third domain refers to teachers' oral ability to engage students in learning and teachers' assessment conduction. The professional aspect demonstrates teachers' activities to better students' learning by reflecting their classwork, communicating with parents, joining the professional community and showing evidence of professional development. These practices can go along with the perception aspects as mentioned in the previous part; i.e. teaching contents, teaching methods and professional development.

English teaching contents vary in different contexts of different purposes and resources. English teaching methods, on the other hand, have undergone three common trends, including traditional approaches before 1960s, classic communicative approaches between 1970s and 1990s, and modern communicative approaches from the late 1990s till now (Richard, 2006, p.6). Plus, the late part of the twentieth century introduces the post-method (Kumaravadivelu, 1994; Richards & Rogers, 2001; Chen, 2014). The oldest approaches prioritise the mastery of grammatical rules, featured by Grammar-translation Method, Direct Method, Audiolingualism (Aural-Oral Method), and Structural-Situational Approach (Situational Language Teaching) (Richard, 2006; Brandle, 2008). The Grammar-translation method focuses on grammar and vocabulary and these language aspects are normally taught deductively. It is derived of developing students' communication in the target language. The Direct Method becomes its opponent, which refutes translation into the mother tongue, but a direct exposure to the target language with oral communication built carefully through teacher-students' exchanges in intensive classes. The language teaching principle evolves to the Audiolingualism, which the presentations of language chunks which are repeated and memorized in its natural context. The Situational Method follows the P-P-P model (presentation-practice-production), in which grammar is taught from the context of a text. However, these methods fall out of fashion because they are hard to have students use language meaningfully and fluently. A focus on separate items of grammar and vocabulary gives the way to a development of communicative competence for communicative purposes like making requests and describing needs, etc. Communicative syllabi are developed with the skill-base and function-base. Nonetheless, classic communicative approaches continue growing till the present. If the classic style is limited to sets of fixed principles, modern communicative teaching, while still placing its emphasis on language users' communicative competence, is more flexible. In reality, teachers may not follow a single method. Or else, they think they are using this method, but in fact their activities illustrate another method. Nonetheless, the diversity in methods are adopted as long as they boost up the student use of language in communication.

3.4. Popular washback models

Alderson & Wall (1993) are accredited as pioneers to build up the first popular washback theory, followed by Hughes (2003), Bailey (1996), Bachman & Palmer (1996), Green (2007) and Shih (2009). Washback aspects pertaining to teachers and their teaching will dominate the discuss room herein, basing on the present research objectives.

In Alderson & Wall's (1993) fifteen-hypothesis framework, eight hypotheses mention the influence of the test on teachers and teaching. A very general statement is claimed first: a test will affect teaching, tailing specific affecting factors embracing teaching contents, methods, rate, sequence, degree and depth of teaching. These authors also state that a test will affect different teachers differently. This is later empirically explained with various washback effects on different teachers in diversified contexts.

Components of washback appear more obviously in models by Hughes (2003), Bailey (1996) and Green (2007). Washback appears in the trichotomy of *participants, process,* and *products*, which "may be affected by the nature of a test" (Hughes, 2003, p.2). The author widens the range of *participants* as language learners and teachers, administrators, materials developers, and publishers, whose perceptions, attitudes, motivations and actions can be impacted by the test. He defines *process* as any of participants' behaviors serving learning goals, including materials development, syllabus design, changes in teaching methods or content, learning and/or test-taking strategies, etc. Finally, *product* covers the learnt contents and their quality.

Three years later, Bailey (1996, p. 264) develops Hughes' (2003) trichotomous model into a washback framework portraying the complicated reciprocal interactions among all the components, commencing from the test and ending in it, too. A new participant as *researchers* is involved; however, "researchers" and "material writers and curriculum designers", compared to "students" and "teachers", are far from direct teaching and learning. Furthermore, the test affects teachers; and teachers, in turns, implement their teaching. In contrast, teachers also exert their impact on the test. This is possibly true in case teachers have the right to make changes with the test, but not true in all situations. In the model, "participants" and "products" enjoy four corresponding labels each. "Process", in other words learning/teaching/designing/researching, enables "participants" to actualize their "products". The question how the process takes place will be of great importance to guide washback researchers; hence, it requires immense elaborations by the followers.

In the same year 1996, Bachman & Palmer (p. 147) provide aspects concerning washback to teachers by questioning the consistence between (i) “the areas of language ability to be measured” and “those that are included in teaching materials”, (ii) “the characteristics of the test and test tasks” and “the characteristics of teaching activities”, (iii) “the purpose of the test” and “the values and goals of teachers and of the instructional program”. Content factors are taken into considerations, i.e. gauged language skills and taught ones, test characteristics, teaching practices. Furthermore, point (iii) in their theory can share several values with Alderson & Wall’s (1993) theory. Stated from this perspective, washback is shown when test characteristics are validated, and there is a link amongst the test content and syllabus content as well as teachers’ beliefs and practices.

In 2009, Shih (p. 199) presented the most detailed washback model of washback to teaching. The advanced aspect of the test is the dynamic convergence of well-listed contextual factors, test factors and teacher factors to impact teaching practices. The author adds letter “t” as a sign of the changing nature of washback over time. According to Shih (2009, p. 200) the italicized factors are significant in her research, while the underlined ones are her new recommendations. The first component is analysed into three layers from the national factors to the school factors and then to the course level. Washback, addressed in this way, can be said to extend its scope to impact as discussed in the part of washback concept above. Test factors, including the test status, test content, test quality, etc. are claimed to directly affect teachers’ inherent factors and teachers’ teaching practices. Teacher factors, illustrated by their educational background and present perceptions of the test status and test quality, etc. will be reflected in their teaching. In this model, teaching aspects are shaped vividly, in comparison to other frameworks, which can be described in terms of teaching contents, teaching methodology, and even psychological features. Great efforts would be needed to specify the information in the box washback of tests on teaching. If Bailey (1996) presents the two-sided impact between the test and the participants, Shih (2009) uses one-sided arrow targeting washback of the test to teaching. Possibly, washback of the test has its significant meaning in giving feedback to the policymakers and the test users, teachers included, from which positive changes may happen.

Overall, the washback mechanisms have developed from Alderson and Wall’s (1993) radical framework stating the influence of the test on teachers in terms of teaching contents, methodology, degree, depth, as well as teachers’ attitude towards these aspects to Shin’s (2009) more complicated model showing various complicated washback facets. The washback degree and scope will vary regarding personal factors and contextual factors, which need detailed studies in differing institutions.

4. Results & Discussion

4. 1. The updated washback framework.

Up to now, six washback models by Alderson & Wall (1993), Bachman & Palmer (1996), Bailey (1996), Green (2007), Hughes (2003), and Shin (2009) have been critically reviewed, with the focus on aspects pertaining to teachers and teaching. The conceptual framework of washback for this research is formulated from the integration of the above mechanisms to investigate washback of a test, EAT, to teachers’ perceptions and practices at a university in Vietnam. Teachers’ perceptions of students' level and motivation have certain impacts on their practices (Chen, 2002; Furaidah et al. 2015).

The conceptual framework is original in the way the components are organised as well as the addition of an element of teachers' professional development. This new factor is significant when it is commonly agreed that teachers play as a driving force in producing positive washback (Bailey, 1996, Liauh, 2011; Tsagari, 2011; Ahmad & Rao, 2012; Onaiba, 2013; Antineskul & Sheveleva, 2015; Spratt, 2005). It is also significant in the Vietnamese context when the National Foreign Language Project emphases improving teachers' capacity, but there is deficiency in researching how teachers perceive the mission and how they practice it. Vu (2016) is one among few authors mentioning assessment as a trigger for professional development (PD), but numerous challenges are found out. For example, teachers found PD was "too difficult", they "don't have enough time", they "haven't got anything that's really of interest"; and they were "not yet confident enough" (p. 123)

In the washback model, the test is the focal part of this study framework with its constructs and characteristics. There is only one participant role of teachers in relation with their characteristics and values which are linked to their perception of the test status as well as test difficulty.

TEST WASHBACK TO TEACHERS		
Perceptions of the test and students	Perceptions of teaching	Teaching practices
1. Test status	1. Teaching contents	1. Teaching contents
2. Test purpose	1.1. Materials	1.1. Materials
3. Test quality	1.2. Skills	1.2. Skills
4. Test resources	2. Teaching methodology	2. Teaching methodology
5. Students' language capacity for the test	2.1. Teaching approach	2.1. Teaching approach
	2.2. Instructional language	2.2. Instructional language
6. Students' attitude and motivation toward learning English in general	2.3. Instructional time	2.3. Instructional time
	2.4. Time allocation in differing skills	2.4. Time allocation in differing skills
7. Students' attitude and motivation toward the test	2.5. In-class assessment	2.5. In-class assessment
	3. Teacher professional development	3. Teacher professional development
	3.1. Teachers' self-training	3.1. Teachers' self-training
	3.2. Teachers' collaboration with on-site colleagues	3.2. Teachers' collaboration with on-site colleagues
	3.3. Teachers' socialization with the external professional community	3.3. Teachers' socialization with the external professional community

Figure 1. A conceptual framework for test washback to teachers' perceptions and practices

The interaction among these factors are very complicated, which can only be shaped in a detailed study. It can be argued that the perceptions and practices teachers reveal can return as determinant factors to improve the quality of the test. However, the current research is restricted to the impact of the test to teachers before the test event.

4.2. Empirical research on washback to EFL teachers' perceptions and practices

4.2.1. Washback of the test to teachers' perceptions of the test and the students

The implementation of the test exerts washback to teachers' perceptions of the test itself and the students who are the test takers in different ways. Generally, high-stakes tests attract more attention than their low-stakes counterparts, and students of better language abilities seem to attract more teachers' investment of test tasks (Shohamy, 1993; Tsagari, 2007; Chen, 2002; Cheng, 2004).

Regarding teachers' knowledge, attitudes and beliefs of the test, negative reports are founded in most of the research (Hughes, 2003; Shohamy, 1993, Tsagari, 2007; Wang, 2010; Mahmoudi; 2013; Onaiba, 2013; Tayeb et al, 2018, etc.). Hughes (1998) in the context of Turkey, reported teachers' negative attitudes toward the test because it caused their tensions and anxiety. Shohamy (1996) found out the same results with an Arabic test in Isareal. Teachers felt degraded when the test did not help them in their future teaching but demand them. As for the test quality, teachers were not consulted before the final test was designed, which led to their feeling of humiliation and the thought of discrepancies between the test quality and the students' ability. Tayeb et al. (2018) from Yemen was in line with Shohamy (1993) when they reported teachers' bad feeling toward the test because "they were passive when test design, administration and evaluation depends on the High committee of examination" (p.454). In terms of the test status, teachers expressed opposing attitudes toward the Arabic exam, but supports for the EFL exam because the latter could effectively back up students' communicative skills and their future. Despite their positive attitude, teachers suffered from an overloaded volume of materials and stress on students' success or failure (Shohamy, 1996). Shalehi & Yunus (2002, cited in Mahmoudi, 2013) intesified this deleterous impact on teachers in the enactment of the Irannian National University Entrance Exam (INUEE) in Iran. Most teachers in Lybia also held negative attitudes toward the test (Onaiba, 2013). They thought the new examination had a low quality. It was not effective in either evaluating students' integrated skills or preparing students' language use in the future. Even though, the test was not in alighment with the curiculum content, which could be called "under-representative". The negative perception of the test and the resistance to change is also reflected by Wang (2010) in his washback research of the revised CET in China. Plus, Alderson & Hamp-Lyon (1996) and they pointed out that a large number of teachers are stressed under the feeling of guilt and frustration when they felt it hard to provide interesting language lessons and helping students get their expected score. Thuy Nhan (2013) investigated the impact of TOEIC® as a university graduation conditions in Vietnam reflected teachers demotivation in the implementation of the test as well as suitability in the context. However, other researcher saw a light scenario. Cheng (2004) from Hong Kong got the evidence of a majority of teachers' positive attitude toward HKCEE albeit to a superficial change in their core beliefs and initial pressure. Saville (2009), Tsagari (2011), Antineskul & Sheveleva (2015) supported Cheng. Antineskul & Sheveleva (2015, p. 7) justified the reason for teachers' good feeling toward the test Business English Cambridge (BEC) because it was "a new course" to try, had "extra pay", "exam's popularity among students" and "relevance to main curriculum". In terms of the test resources, globally-recognised tests like TOEFL or IELTS enjoyed rich resources (Wall & Horák, 2006; Read and Hayes, 2003; Shohamy, 1998; Saif, 2006; Tsagari, 2011; Peña Jaenes, 2017) while other test types, especially achievement tests were not of this benefit. Possibly, this would caused difficulties to teachers to get a better insight of the test to have more appropriate actions. Plus, Antineskul & Sheveleva (2015) showed teachers believed the test could enhance their professional development opportunity students' future job opportunity and university's reputation. In terms of their perception of teaching BEC, teachers' exam preparations generate advantages.

"Still, even though their reactions might be "no difference in teaching", the respondents unanimously admit that BEC preparation has a positive influence on their teaching. The particular semantics of explanation include "goal-orientation of BEC exam", "individual approach to teaching English", "preparation of teaching materials", "teachers' (improved) knowledge of content language", "variety of teaching materials and tasks for students", "more time-efficient teaching", "ability to motivate students with the content", "justifying students' tuition costs", "developing teaching techniques for exam skills"Antineskul & Sheveleva (2015, p. 9).

Both negative and positive perceptions are also repored by Mahmoudi (2013) with INUEE although the former overwhelmes the latter. Two in six informants thought the test to be able to test universversity applicants' academic language knowledge while the remanining douted its evaluation value of true language proficiency but route-learning. Li's (1990, cited in, Tsagari, 2007) study on washback of NMET to teaching pointed out that in terms of perceptions, teachers had a negative feeling toward the new test first, but changed to a positive attitude in a few years later and thought it could assist their methodology reform at the classroom level. Bui (2016) discussed teachers' perception of English language teaching in high school in Vietnam under the influence of the Vietnam's College English Entrance Exam (VCEE) merely as one part of her research on that test usefulness. Teachers are asked on two themes: the test difficulty and quality. Most respondents thought the test challenged average students, but it covered the textbook contents. By contrast, less than one fifth of them frowned upon the ability to measure test takers' communication skills, but reading skills, grammatical and lexical knowledge. The quality of the test was relatively fine.

Teachers' perceptions of students' language level can guide their practices. Chen (2002) ffound that if teachers believed students possessed a better ability, they would invest more in them to help them reach the target. By contrast, Furaidah et al. (2015) revealed that teachers' perceptions of students of lower-level would entail their more drilling activities and less communicative ones.

The review of teachers' perceptions of the test and the students uncovers differing, even contrasting research outcomes with justifications in unique contexts. The test status and quality, students' language ability have been mentioned. Nonetheless, other important factors like the availability of test resources, students' motivations towards learning English and towards the test have not reached. These gaps open future research on washback.

4.2.2. Washback of the test to teachers' perceptions of teaching

This part would share simlilar features with the part reviewing teachers' practices in different ELT contexts. However, they are distinctive in such a way that these aspects are reported from researchers' interviews and questionnaires delivered to teachers to reveal their beliefs about teaching, not their actions serving teaching, which would be better investigated in observations, and questionnaires, interviews as well.

The perceptions of teaching are found diversified in terms of teachers' academic and professional experience (Mahmoudi, 2013; Lam, 1994; Shohamy, 1996; Watanabe's (2004). The more experienced teachers are, the more test-oriented their perceptions of teaching are.

On answering the question on teaching contents, teachers believed that if the textbooks have the contents in alignment with the test tasks, they preferred to use them (Onaiba, 2013; Tran, 2016). Other respondents perceived the inadequacy in the textbooks resources; hence, supplement materials were exploited. Li (1990, cited in, Tsagari, 2007) found teachers shifted from linguistic knowledge to communicative ability, supported by imported authentic textbooks and reading materials, which could be a good sign for students' language ability development. Bui (2016) saw the test covered the textbook contents, but teachers used more grammar and vocabulary exercises and less reading, speaking, writing and listening and phonology exercises for students. Negative washback revealed in more linguistic knowledge being focused while less language skills were drilled. Sadighi et al. (2018) found out that teachers believed the textbook used for the university entrance exam in Iran would generate beneficial productions of language, but supplementary materials to prepare students for the test would be used more. They percieved that the syntactical and lexical points in the textbook needed modifying because they were demotivating. They also questioned the textbook which followed the test goal but lacked communicative effects.

With regards teacher' beliefs of teaching methodology, diversities were also reported. Sadighi et al. (2018) found out that teachers spent most of their class time instructing students tasks which were relevant to the test. More time was devoted to grammar explanation and practices. Onaiba (2013, p. 246) showed teachers' acceptance of the new exam tasks to alter their teaching methods. Wall & Horák (2006) questioned the reliability of classroom assessment. Simulation tests are employed but the test condition is not standard. Due to less communicative ability in English is required, teachers perceived that Vietnamese as a means of instruction was more popular in language classes (Bui, 2016). Mahmoudi (2013) found out that experienced teachers owned the perception that INUEE-related tests and materials should account for more class time. Experienced teachers were more exam-oriented than less eperienced counterparts (Lam (1994; Shohamy, 1996).

Washback to teachers' aspects of teaching is not limited to teaching contents and teaching methodology, but is extended to teachers' professional development, because this can contribute to their perceptions and actions. There exists a big gap of the issue in the reviewed research, except few words voiced by Wall & Horák (2006), Thuy Nhan

(2013), and Antineskul & Sheveleva (2015). While Antineskul & Sheveleva (2015) showed teachers believed the test could enhance their professional development, Thuy Nhan (2013) showed teachers' demotivation in their profession due to the test. She linked teachers' professional reputation to the rate of successful/failed students in the exam. A majority of teachers were under pressure when they lacked professional training opportunities of the test but they were assigned to teach how their students could achieve a target score. Wall & Horák (2006) was in line with her.

The above discussion depicts teachers' perceptions of teaching in terms of three aspects: contents, methodology and carreer intensification. Teachers believed in direct teaching to the test with aligning materials, skills and teaching techniques. Nevertheless, many issues in the present study framework has not been reached fully, especially in the area of teachers' development.

4.2.3. Washback of the test to teachers' practices

A huge volume of research has been conducted on the impact of the test on teachers' practices, releasing rich sources of information of both positive and negative washback.

Researchers found in most cases textbooks reflected the (previous) exam formats; hence, teachers use them (Wall & Horak, 1996; Onaiba, 2013; Wantanabe, 2004; Barnes, 2016). For example, Hamp-Lyons (1997) saw the persent textbooks consisted of test-taking strategies and mastery of language grammar, vocabulary and other linguistic features of the previous TOEFLs. Wall & Horak (1996) echoed Hamp-Lyon' view about TOEFL materials. This led to a situation in which Lam (1994) raised the problem of teacher exam/textbook slaves. It may bebe better if authentic materials are brought into the classroom, rather than such commercial materials. Barnes (2016, p. 170) cited Widdowson's (1981) distinction between 'goal- oriented courses' and 'process-oriented courses' to argue that the TOEFL test preparation course features the first type; therefore such teachers' practices were quite justifiable. The instrumental textbooks emerge the closely neat cooperation between test designers and textbook developers to make sure the communicative constructs. On the other hand, these materials should be adapted to the contextual factors like students' needs and students' level. Furaidah et al. (2015) investigated washback to non-like ENE (English national examination) materials and like-ENE materials with the shift to the former in the later part of the course, especially for better students. Students' level affected teachers' choice of materials. Low-level students benefited from materials within their capacities. Saif (2006) had evidence of teachers' alignment beween materials and students' needs as well as test purpose. Supplement materials are used to compensate for inadequate linguistic knowlege, i.e, grammar and vocabulary, in the text book (Wall & Horák, 2006). Anderson, Wall and Hamp-Lyon (1993, 1996) showed that teachers narrowed down or abandon their textbooks for intensive work with past papers and commercial publications to prepare their students for the exam. Like Alderson & Wall (1993, 2005), Tsagari (2011), Onaiba (2013), Salehi, Unus & Salehi (2012) Saif, (2006), Read and Hayes (2003), Cheng (2004) saw teachers change their teaching contents under the influence of the test. This happens to tests which have rich and systematic sources like TOEFL or IELTS (Wall & Horák, 2006; Read and Hayes, 2003; Shohamy, 1993; Saif, 2006; Tsagari, 2011; Peña Jaenes, 2017). Teachers contrentrated on manifesting the test tasks due to of time limitation and workload. Skills to teach were narrowed down to those to be tested in the test. For example, teachers narrowed down the learning skills by focusing on listening and reading (Furaidah et al., 2015). Or, listening and writing which are not tested, are excluded in teaching (Rahman, 2014). Wall & Horák (2006) observed classrooms, seeing that teachers realized listening strategies but did not breakdown them into teachable skills (p.108). Reading was used as homework. Grammar and vocabulary were taught in points. That means linguistic knowledge is prioritized to communicative skills. In Vietnam, Thúy Lan Nguyễn (2017) conducted a study on washback of VSTEP , reporting positive washback from the exam. It classified students of different language skills at different levels, which acted as a guideline to design the curriculum.

More specifically about methodology, most classed are reported teacher-centered with traditional methods. (Alderson & Wall, 1993; Wang, 2010; Onaiba, 2013; Tsagari, 2011; Taqizadeh & Birjandi, 2015). Barnes (2016) revealed the dominant role of teachers in class, the heavy reliance on commercial TOEFL materials and the limited class activities. It is interesting to see Furaidah et al. (2015, p.49) pointed two types of teaching, including regular teaching (students-oriented) and drilling teaching (teacher-oriented/ test-oriented for different phase to approach the final test. The classroom interaction in the second phase was described as the teacher questioning-student answering time. Wang (2010) in his washback research of the revised CET. There are gaps between the information collected from the survey and that from the observations. For examples, while the interview reported the positive shift in methodology, teachers' actions in class were still teacher-centred. Although new books designed communicative activities in speaking lessons, teachers still skipped them and maintained the knowledge transformation mode like in the old version. While 69% of teachers perceived the detrimental impact of tests on teacher, 65% considered CET a student's learning driving force (Wang, 2010, p. 190). A number of contradictions in teachers' perceptions and actions have been pinpointed in the research, which means that washback is a too complicated matter, and data triangulation is necessary. He concluded that the revised CET did almost no change in teachers' beliefs and practices (Wang, 2010, p. 215). Thuy Nhan (2013) saw teachers increase their class involvement to both attaining curriculum goals and supporting students' achieving the TOEIC® certification. They had more power in their language class and make more careful decisions concerning teaching and assessment. Thúy Lan Nguyễn (2017a) reported significant changes in pedagogical class activities when teachers shifted students to active learning via project learning and integrative learning. Specifically, testing and assessment in light of VSTEP varied its forms, emphasizing on continuous assessment. Also concerning assessment, Onaiba (2013) reported the use of mock exams, especially when the exam comes closer. Time allocation during course is also a matter of concern. Furaidah et al. (2015) reported time division was decided by the policy makers with more time for the later part of the course. However, it is important to know how teachers themselves spend their own time budget in their own class. No other other research paid attention to this while different time allocation on different parts of study can results in different washback intensity level. In addition, L1 as instructional language was popular in language classes (Onaiba, 2013; Bui; 2016; Salehi, Unus & Salehi, 2012). The reasons for using mother tongue were to save time and to address low level students.

Now it comes to the issue of professional development with little information revealed, except in Thuy Nhan (2003), Wall & Horák (2006) and Antineskul & Sheveleva (2015). Although Antineskul & Sheveleva (2015) showed teachers believed the test could enhance their professional development, they lack the detailed descriptions of the activities. Thuy Nhan (2003) and Wall & Horák (2006) complained very limited professional development opportunities for teachers at the studied universities. More evidence should be found on this to see how teachers develop their career themselves and in the communities, from which policy makers can have appropriate actions to build up their employeees' capacity.

The documentation of washback of English tests to teachers' practices have called out the current situations of the impact of the test to teaching, as well as pointed out research gaps. Teaching contents tend to be twisted with the test contents and test format, but teaching methodology seems to be more stative. Traditional methods are more frequently reported than communicative methods. This is in line with Cheng's (1997) remark that washback is more dominant in teaching materials than methodology. Once again, the matter of teachers' time allotment or teachers' professional development has been left less touched.

In conclusion, two major flaws of previous research papers can be enumerated as follows. First, washback to teachers' perceptions has been discussed widely; nonetheless, little information about the way teachers perceive their students' language ability, motivation and their actions on professional development and the way they give instructions in class and practice professional development. Thirdly, washback studies tend to concern highly-recognised exams, IELTS, TOEFL or EUUs, for example, while leaving smaller-scale tests like achievement ones nearly untouched. Further research are hoped to fill into the research gap.

5. Conclusions

In conclusion, the current research has managed to scaffold a new theoretical framework of the test washback to teacher perceptions and practices. Popular washback theories suggested by Alderson and Wall (1993), Hughes (2003); Bailey (1996), Bachman & Palmer (1996), Green (2007) and Shih (2009) have been analysed, focusing on teacher factors. The updated framework are composed of three aspects, namely teachers' perceptions of the test and the students, teachers' perceptions of teaching and teachers' practices. A subsequent review of empirical research on washback in various contexts reveals several research gaps for further research agenda. Test validity can be studied as a base for the washback research. Teacher professional development under the test impact has been largely researched. Plus, more washback research focuses on high-stake tests, leaving low stake ones little touched, too.

References

Alderson, J.C., & Wall, D. (1993). Does washback exist? *Applied Linguistics*, *14*, 115-129.

Alderson, J. C. & Hughes, A. (1981). *Issues in language testing ELT documents 111*.

Alderson, J. C., & Banerjee, J. (2002). Language testing and assessment (Part 2). *Language Teaching*, *35*(02), 79–113. https://doi.org/10.1017/S0261444802001751

Alderson, J. C., & Banerjee, J. (2001). Language testing and assessment. *Language Teaching*, *34*(Part I), 213–236. https://doi.org/10.1017/S0261444801001707

Alderson, J.C., & Wall, D. (1993). Does washback exist? *Applied Linguistics*, 14, pp. 115-129.

Ahmad, S., & Rao, C. (2012). Examination washback effect : syllabus, teaching methodology and the learners 'Communicative Competence. *Journal of Education and Practice*, *3*(15), pp.173–183.

Antineskul, O., & Sheveleva, M. (2015). *Teachers' perceptions towards BEC exams in Russia: A qualitative study*. Shevelava.

Bachman, L. F., Palmer, A. S. (1996). Language testing in practice: designing and developing useful language tests. *Oxford Applied Linguistics*. https://doi.org/10.2307/328718

Barnes, M. M. (2016a). Washback and CLT. *Centre for Applied Linguistics Research*, *7*(2).

Barnes, M. M. (2016b). The washback of the TOEFL iBT on English language programs in Vietnam. *Australian Journal of Teacher Education*, *41*(7), 246. https://doi.org/10.14221/ajte.2016v41n7.10

Barnes, M. M. (2017). Washback: exploring what constitutes "good" teaching practices. *Journal of English for Academic Purposes*, *30*, 1–12. https://doi.org/10.1016/j.jeap.2017.10.003

Bailey, K. M. (1996). *Washback in language testing* e*ducation journal*. New Jersey: Educational Testing Service, Princeton, New Jersey, RM-99-4. https://doi.org/10.11648/j.edu.20150401.12

Borg, S. (2003). Teacher cognition in language teaching : A review of research on what languae teachers think, know, believe , and do. *Language Teaching*, *36*(2), 81–109. https://doi.org/10.1017/S0261444803001903

Borg, S. (2006). Teacher cognition and language education. *Research and practice.* London, UK: Continuum.

Bui, T. H. (2018). *Washback of the proficiency assessment as exit requirement on non-English major students at a law school* (Masters' Thesis). Retrieved from http://repository.ulis.vnu.edu.vn/bitstream/ULIS_123456789/1741/1/Bui%20Thanh%20Huong.pdf

Bui, T. S. (2016). *The test usefulness of the Vietnam's college English Entrance Exam* (Master's Thesis). Korea University, Seoul.

Bullock, N. (2017). *Learning & testing alignment: towards positive washback*. Powerpoint presentation at the ICAO LPRS - 10 years on: Progress or Pain? Dubrovnik.

Chen, L.M.D. (2002). *Taiwanese junior high school English teachers' perceptions of the washback effect of the basic competence test in English* (Unpublished Doctoral Dissertation). The Ohio Satate University, Ohio.

Chen, M. (2014). Postmethod pedagogy and its influence on EFL teaching strategies. *English Language Teaching*, *7*(5), pp. 17–25. https://doi.org/10.5539/elt.v7n5p17

Cheng, L, Watanabe, Y & Curtis, A. (2004) (Eds). *Washback in Language Testing: Research Contexts and Methods*. New Jersey: Lawrence Erlbeaum Associates, Publishers. https://doi.org/10.4324/9781410609731

Cheng, L., Andrews, S., & Yu, Y. (2011). Impact and consequences of school-based assessment (SBA): Students' and parents' views of SBA in Hong Kong. *Language Testing*, *7*(2), 221–249. https://doi.org/10.1177/0265532210384253

Cheng, L. (1999). Changing assessment: Washback on teacher perceptions and actions. *Teaching and Teacher Education*, *15*(3), 253–271. https://doi.org/10.1016/S0742-051X(98)00046-8

Cheng, L. (2008). Washback, Impact and Consequences. *Encyclopedia of Language and Education: Language Testing and Assessment*, *(2nd Ed.) (7),* 349–364. https://doi.org/10.1007/978-0-387-30424-3

Cheng, L. & Curtis, A. (2012). Test impact and washback: Implications for teaching and learning. In S. Coombe, Christine; Davidson, Peter; O'Sullivan, Barry; Stoynoff (Eds.), *The Cambridge Guide to Second Language Assessment*. New York: Cambridge University Press.

Cheng. L. (2014). The washback effect of a pbublic examination change on teacher's perceptions toward their classroom teaching. In C., Liying, W., Yoshinori., A., Curtis (Eds.). *Washback in Language Testing: Research Context and Method* (pp.19-26). Marwah, New Jersey: Lawrance Erlbaum Associates, Inc. Publishers.

Cheng, L., Sun, Y. & Ma, J. (2015). Review of washback research literature within Kane's argument-based validation framework. *Language Teaching*, *48*(4), 436–470. https://doi.org/10.1017/S0261444815000233

Danielson, C. (1996). *Enhancing professional practice: A framework for teaching*. Alexandria, VA: Association for Supervision and Curriculum Development.

Fives, H., Gill, M. G. (2014). *International handbook of research on teachers' beliefs*. London: Taylor & Francis Ltd.

Furaidah, S., A., & Widiati, U. (2015). Washback of English national examination in the Indonesian context Utami Widiati. *Teflnjournal*, *26*(1), pp. 36–58. https://doi.org/10.15639/teflinjournal.v26i1/36-58

Green, A. (2007). IELTS Washback in Context: Preparation for academic writing in higher education. *Studies in Language Testing 25*. Cambridge: Cambridge University Press.

Green, A. (2013). Washback in language assessment. *International Journal of English Studies*, *13*(2), 39–51.

Hamp-Lyons, L. (1997). Washback, Impact and Validity: Ethical Concerns. *Language Testing*, 14(3), 295-303. Hsieh, C. (2017). The Case of Taiwan : Perceptions of College Students About the Use of the TOEIC ® Tests as a Condition of Graduation. In *ETS Research Report Series* (pp. 1–12).

Hsu, H. F. (2009). *The impact of implementing English proficiency tests as a graduation requirement at Taiwanese universities of technology* (Doctoral Dissertation) Retrieved from http://etheses.whiterose.ac.uk/576/

Hughes, A. (2003). Testing for Language Teachers. *Australian Review of Applied Linguistics* (Vol. 27). https://doi.org/10.1017/CBO9780511732980

Kumaravadivelu, B. (1994). The postmethod condition: Emerging strategies for second/foreign language teaching [J]. *TESOL Quarterly*, 28, 27-47.

Lam, H P (1994) Methodology washback--an insider's view. In Nunan, D., Berry, R. and Berry, V. (Eds) *Bringing about change in language education: Proceedings of the International Language in Education Conference* (pp. 83-120) Hong Kong: University of Hong Kong.

Latimer, D. G. (2009). *Washback effects of the Cambridge Preliminary English Test at an Argentinean bilingual school*. Unpublished paper.

Le, T.H.D. (2017). *Exploring teachers' perceptions and practices of English for Specific Purposed (ESP) in Vietnames context* (Doctoral Dissertation). Retrieved from ULIS Doctoral Dessertations.

Le, V. C. (2011). *Form-Focus instruction: A case study of Vietnamese teachers' beliefs and practices* (Doctoral Disseration). Retrieved from https://researchcommons.waikato.ac.nz/bitstream/handle/10289/5253/thesis.pdf;jsessionid=159888BF75888614104738591EE443CF?sequence=3

Lemmetti, J. (2014). *What makes a good language test in EFL*? Gôteborg Universite.

Liauh, Y.-H. E. (2011). *A study of the perceptions of English faculty and students of Exit English Examinations at Taiwan's technological and vocational higher education institutions* (Doctoral Dessertation). Retrieved from http://search.proquest.com/docview/874150865?accountid=14548%5Cnhttp://metadata.lib.hku.hk/hku?url_ver=Z39.88-2004&rft_val_fmt=info:ofi/fmt:kev:mtx:dissertation&genre=dissertations+%26+thes

Mahmoudi, L. (2013). Iranian Pre-university English Teachers ' Perceptions and Attitudes towards the Iranian National University Entrance Exam : A Washback Study. *International Journal of Education & Literacy Studies*, *1*(2), 309–310. https://doi.org/10.7575/aiac.ijels.v.1n.2p.47

Manjarrés, N. B. (2004). Washback of the foreign language test of the state examinations in Columbia: A case study. *Arizona Working Papers in SLAT*, *12*, 1–19.

McNamara, T. (2000). *Language Testing*. Oxford, UK: Oxford University Press.

Messick, S. (1996). Validity and washback in langauge testing. *Language Testing*. Retrieved from http://ltj.sagepub.com/content/13/3/241.short

Nguyen, P.N., Griffin, P. &, & Izard, J. (2008). *Toward beneficial washback effects of a test*, 1–27.

Nguyen, P. N. (1997). *Washback effects of international English language testing system at the Vietnam National University* (Doctoral Dissertation) University of Melbourne, Oppenheim.

Nguyen, Thị Linh. (2017). Impacts of Vietnamese standardised test of English proficiency on the first year students' English language learning. *Đề án ngoại ngữ quốc gia 2020*. Retrieved from https://dean2020.edu.vn/vi/news/Tin-tuc/impacts-of-vietnamese-standardized-test-of-english-proficiency-vstep-on-the-first-year-students-english-language-learning-410.html

Nguyen, T. L. (2017). The Effects of VSTEP on Students' Learning Motivation. *Kỷ yếu hội thảo khoa học quốc gia dành cho học viên cao học và nghiên cứu sinh lần thứ nhất (GRS) 2017*. Graduate Research Symposium (GRS) 2017.

examination on teachers ' instructional practices , materials and curriculum. Leicester: University of Leicester.

Onaiba, A. M. E. (2013). *Investigating the washback effect of a revised EFL public examination on teachers ' instructional practices , materials and curriculum*. Leicester: University of Leicester.

Pajares, M. F. (1992). Teachers- beliefs and educational research: Cleaning up a messy construct. *Review of Educational Research*, 62(3), 307-332.

Pan, Y. (2009). A review of washback and its pedagogical implications. *Journal of Science, Foreign Languages*, *25* (February), 257–263.

Pan, Y. C., & Newfields, T. (2012). Tertiary EFL proficiency graduation requirements in Taiwan: A study of washback on learning. *Electronic Journal of Foreign Language Teaching*, *9*(1), 108–122.

Papakammenou, I. (2013). *Examining Washback in Multi-exam Preparation Classes in Greece (A Focus on Teachers' Teaching practices)*.pdf. EALTA Conference Turkey.

Peña Jaenes, V. (2017). Testing Writing: the Washback Effect on English Courses. *The Grove - Working Papers on English Studies* (Vol. 24). Jaén, Andalucía, Spain.

Pierce, B. N. (1992). Demystifying the TOEFL reading test. *TESOL Quarterly.* 26(4), pp. 665-691.

Rahman, S. (2014). *Teachers Suffering from Exam Washback Effect: Exclusion of Practicing Speaking and Listening Skills in English Classes*. Dhaka: BRAC University.

Read, J., & Hayes, B. (2003). *The Impact of IELTS on Preparation for Academic Study in New Zealand*. *IELTS Research Reports* (Vol. 4). Canberra. Retrieved from https://search.informit.com.au/documentSummary;dn=909013632781357;res=IELHSS

Richard, C. J. (2006). *Communicative Language Teaching Today*. New York: Cambridge University Press.

Richards, J. C., & Lockhart, C. (2007). *Reflective Teaching in Second Language Classrooms*. *The Modern Language Journal* (Vol. 79). New York: Cambridge University Press. https://doi.org/10.2307/329404

Sadighi, S., Yamini, M., & Bagheri, M. S. (2018). Investigating preuniversity EFL teachers ' perceived wash-back effects of university entrance exams and teaching materials on students ' learning objectives and teachers ' class performance. *Cogent Social Sciences*, *00*(00), 1–16. https://doi.org/10.1080/23311886.2018.1546448

Saif, S. (2006). Aiming for positive washback: a case study of international teaching assistants. *Language Testing*, *23*(1), 1–34. https://doi.org/10.1191/0265532206lt322oa

Saville, N. D. (2009). *Developing a model for investigating the impact of language assessment within educational contexts by a public examination provider*. Bedfordshire: University of Bedfordshire.

Salehi, H., Yunus, M. M., & Salehi, Z. (2012). Teachers' perceptions of high-stakes tests: a washback study. *International Journal of Social Science and Humanity (IJSSH)*, *2*(1), 70–74.

Shih, C.-M. (2007). A new washback model of students' learning. *Canadian Modern Language Review*, *64*(1), 135–161. https://doi.org/10.3138/cmlr.64.1.135

Shih, C.-M. (2009). How tests change teaching: A model for reference. *English Teaching: Practice and Critique*, *8*(2), 188–206.

Shohamy, E., Or, I. G., & May, S. (2017). *Language testing and assessment*. Switzeland: Springer. https://doi.org/10.1007/978-319-02261-1

Shohamy, E. (1993). The power of tests : The impact of language tests on teaching and learning. NFLC occasional paper. *The National Foreign Language Center*, 1–20. Retrieved from https://eric.ed.gov/?id=ED362040

Shohamy, E. (1998). Critical language testing and beyond. *Studies in Educational Evaluation*, *24*(4), 331–345. https://doi.org/10.1016/S0191-491X(98)00020-0

allr.ir/index.php/JALLR/article/view/63

Tayeb, Y., Abd Aziz, M., Ismail, K., & Khan, A. (2014). The Washback Effect of General of the General Secondary English Examinations (GSEE) on teaching and Learning. *GEMA Online Journal of Language Studies*, *14*(3), 83–103.

Tayeb, Y. A., Abd Aziz, M. S., Ismail, K. (2018). Predominant Washback of the General Secondary English Examination on Predominant Washback of the General Secondary English Examination on Teachers. *International Journal of Engineering and Technology*, *7*(August). https://doi.org/10.14419/ijet.v7i3.21.17211

Taylor, L. (2005). Washback and impact. *ELT Journal*, *59*(2), 154–155. https://doi.org/10.1093/eltj/cci030

Thuy Nhan. (2013). The TOEIC ® Test as an Exit Requirement in Universities and Colleges in Danang City , Vietnam : Challenges and Impacts. *International Journal of Innovative Interdisciplineary Research*, *2*(June), 33–50.

Tran, T. D. (2016). Impact of using TOEIC as an exit requirement at a public university in Vietnam. In *4th British Council New Directions in English Language Assessment: Standardised Testing and Proficiecy Scales*. Retrieved from https://www.britishcouncil.vn/sites/default/files/new_directions_2016_dr_duyen_tran_impact_of_using_toeic_as_an_exit_requirement_at_a_public_university_in_vietnam.pdf

Tsagari, D. (2007). *Review of washback in language testing : How has been done ? What more needs doing*? Lancaste: Lancaster University, UK.

Tsagari, D. (2011). "Washback of a high-stakes English exam on teachers'' perceptions and practices"." *Selected Papers from the 19th ISTAL*, (1996), 431–445.

Vu, T. P. A. (2016). 25 years of language assessment in Vietnam : Looking back and looking forward. In *New Directions in English Language Assessment in Vietnam*. Retrieved from https://www.britishcouncil.vn/.../new_directions_2016_dr_vu_thi_phu...

Wall, D., & Horák, T. (2006). *The impact of Changes in the TOEFL Examination on Teaching and Learning in Central and Eastern Europe: Phase 1, The Baseline Study*. Retrieved from https://www.ets.org/Media/Research/pdf/RR-06-18.pdf

Wang, J. (2010). *A study of the role of the "teacher factor" in washback* (Dooctoral disseratation). Retrieved from http://digitool.library.mcgill.ca/webclient/StreamGate?folder_id=0&dvs=1565578944761~588

Watanabe, Y. (2004). Methodology in Washback Studies. In C., Liying, W., Yoshinori & A., Curtis (Eds.), *Washback in Language Testing: Research Context and Method* (pp.19-26). Marwah, New Jersey: Lawrance Erlbaum Associates, Inc. Publishers.

Wenyuan, Z. (2017). The washback effect of CET spoken English test upon college English teaching. *Canadian Social Science*, *13*(1), 62–68. https://doi.org/10.3968/9241

Zeng, H. (2015). Teacher beliefs as a complex system: English language teachers in China. NY: Springer.

KHẢO CỨU VỀ CÁC MÔ HÌNH VÀ NGHIÊN CỨU THỰC NGHIỆM Ở VIỆT NAM VÀ TRÊN THẾ GIỚI VỀ TÁC ĐỘNG DỘI NGƯỢC CỦA BÀI THI TỚI GIÁO VIÊN DẠY TIẾNG

Đinh Minh Thu

Đại học Hải Phòng
171 Phan Đăng Lưu, Kiến An, Hải Phòng, Việt Nam

Tóm tắt: Tác động dội ngược hay ảnh hưởng của bài thi tới việc dạy và học đã trở thành chủ đề nghiên cứu hấp dẫn ở lĩnh vực kiểm tra và đánh giá ngôn ngữ trong vòng hai mươi năm trở lại đây do ứng dụng quan trọng của những nghiên cứu này giúp các nhà hoạch định chính sách và những người thực thi chính sách đảm bảo đánh giá có tính chính xác và công bằng. Hiện nay, những nghiên cứu này được nhiều nhà nghiên cứu ở Việt Nam quan tâm hơn do yêu cầu cải tiến công tác kiểm tra đánh giá được Đề án Ngoại ngữ Quốc gia 2020 (kéo dài đến 2025) đặt ra. Tác động dội ngược diễn ra một các tích cực hoặc tiêu cực tùy theo việc nó đẩy mạnh hay ngăn cản hiệu quả học tập. Giáo viên được coi là nhân tố tiên phong trong cơ chế tác động. Chỉ có một mô hình về tác động dội ngược tới giáo viên được Shih (2009) đề xuất. Bài viết này mục đích khảo cứu các mô hình tác động dội ngược, bao gồm mô hình của Shih (2009), để xây dựng lên một mô hình tác động dội ngược của bài thi tới nhận thức và hành động của giáo viên. Mô hình mới này sẽ được sử dụng để khảo cứu các nghiên cứu thực nghiệm trước đây về tác động dội ngược trong và ngoài Việt Nam giúp tìm ra những điểm các nhà nghiên cứu đã tìm ra và cả những điểm chưa được nghiên cứu. Phương pháp nghiên cứu trong bài viết này là định tính thông qua phân tích hơn bốn mươi tài liệu gồm nhiều thể loại như sách, luận văn, luận án, bài báo, v.v. Nội dung thảo luận chính của bài viết gồm hai phần: khảo cứu các mô hình lý thuyết và khảo cứu các nghiên cứu thực nghiệm. Kết quả khảo cứu chỉ ra rằng tác động dội ngược của bài thi tới nhận thức của giáo viên liên quan tới nhận thức của giáo viên về bài thi, về năng lực ngôn ngữ của người học, về nội dung và phương pháp giảng dạy, về việc giáo viên phát triển chuyên môn. Thêm vào đó, tác động dội ngược của bài thi lên hành động của giáo viên bao gồm việc họ lựa chọn nội dung giảng dạy, phương pháp giảng dạy và phát triển chuyên môn. Yếu tố phát triển chuyên môn là yếu tố mới trong khung nghiên cứu mới tạo ra. Điều này có ý nghĩa quan trọng làm tăng nhận thức của giáo viên về việc phát triển chuyên môn. Bài viết này mong muốn đóng góp cho việc làm rõ bức tranh về nghiên cứu tác động dội ngược của bài thi giúp các nhà nghiên cứu, giáo viên và quản lý có quan tâm đến lĩnh vực này hiểu rõ hơn tình hình nghiên cứu hiện nay.

Từ khóa: tác động dội ngược, mô hình tác động dội ngược, bài kiểm tra ngôn ngữ, nhận thức của giáo viên, hành động của giáo viên

LEVELS OF ANALYSIS FROM THEORY TO PRACTICE - CASE STUDY: U.S. POLICY IN THE EAST SEA UNDER THE TRUMP ADMINISTRATION

Pham Minh Thu*

Diplomatic Academy of Viet Nam, Americas Department, Ministry of Foreign Affairs
69 Chua Lang, Dong Da, Hanoi, Vietnam

Received 02 March 2020
Revised 15 May 2020; Accepted 27 May 2020

Abstract: Three levels of analysis (individuals, nation states and international system) which have been widely recognized in foreign policy are applicable in explaining whether the Trump Administration has actually had a policy in the East Sea. At individual level, President Trump first announced the U.S. "Free and Open Indo – Pacific Strategy" in November 2017, in which the East Sea is known as an important factor of security pillar to preserve vitality of free and open air and maritime space in the region. It was then echoed at national level by national strategies, including the National Security Strategy, National Defense Strategy and National Military Strategy. Also, members of the U.S. Cabinet and Congress have shown support for the Indo – Pacific Strategy in general and the U.S. policy in the East Sea in particular. In order to cope up with China's assertiveness in this waterway at systematic level, the Trump Administration has unceasingly projected considerable its power, including increasing military presence and regional maritime capacity building efforts as well as frequent freedom of navigations (FONOPS). In the context that territorial disputes remain complex and unforeseeable, it is valuable to have a thorough look at the Trump's East Sea policy for claimants in this water, including Viet Nam.

Keywords: Levels of analysis, Trump, policy, East Sea.

1. Introduction

Three levels of analysis (individuals, nation states and international system) have been long developed and widely recognized in foreign policy. Under the Trump Administration, these levels are applicable in explaining whether the Administration has set a policy in the East Sea. In the context that territorial disputes in this region remain complicating, it is worth seeking answer for this research question which will have strategic implications for claimants, including Viet Nam.

* Tel.: 84-912610000
Email: phamminhthu251@gmail.com

2. Levels of analysis

It is disputable whether the U.S. Administrations, including the Trump Administration, have ever had an overall East Sea policy. If yes, how it has been formatted and implemented to achieve its national interests in this waterway. Hence, the article is tracking down the three levels of analysis to seek answers for these above-mentioned research questions.

In the book "Man, the State and the War" published in 1959, Kenneth Waltz first mentioned three analytical levels (which are also known as "images"), including individuals, nation states and international system. According to Kenneth Waltz, there are three "images" to explain conflicts and wars in international relations. While individual level points out human nature such as ambitions, selfishness and characteristics have impacts on foreign policy making, national state level helps explain that domestic politics are causes of wars. Otherwise, international level helps define forces which put constraint on individuals and states in policy making (Ikenberry, 2014) . Kenneth Waltz assumed that international level outperforms other levels.

These levels were then developed by David Singer (1961) in the article "The level of Analysis Problem in International Relations" in World Politics. Singer (1961: 77-92) regarded international level as "the most comprehensive of the levels available, encompassing the totality of interactions which take place within the system and its environment". However, Singer also pointed out its shortcomings. Particularly, this level "exaggerates the impact of the system upon the national actors and, conversely, discounts the impact of the actors on the system". Meanwhile, national level "permits significant differentiation among our actors in the international system". Additionally, Singer explained that nations include groups of individuals in a certain institution. Then, it is necessary to study individual role in the policy making.

Modern scholars have shown their interest in the three levels of analysis, namely Professor J.T. Rourkev and M.A. Boyer, Connecticut University (Rourke & Boyer, 2010) or Christophe Barbier, Norwhich University (Academia) to explain policy-making and implementing process. Unlike other scholars, Christophe Barbier assumed that individual level plays a core role in the policy making.

Generally speaking, levels of analysis are popularly recognized in foreign policy even though there may be different views on which level overwhelm others. Remarkably, the three levels are not independent. Instead, they are interactive in the process of foreign policy making and implementing, depending on certain historic periods and specific matters.

3. U.S. policy in the East Sea under the Trump Administration

Levels of analysis and their interactions are basically explainable in the US politics in general and its policy making in particular. Through a profound study of the three analytical levels, an East Sea policy under the Trump Administration should be revealed.

First and foremost, at individual level, U.S. President leads the formation of foreign and domestic policies in the U.S. Administration as described in Article 2 of the U.S. Constitution. Then, is is unexceptional for President Trump who is well-known for his strong leadship. Under the Trump's Administration, "American First" approach does not stop him from attaching global strategic focus to the Indo – Pacific with the aim to maintaining the U.S.'s super power status. After taking office in January 2017, President Trump early eliminated "Rebalancing" and replaced it with a new "Free and Open Indo – Pacific Strategy". At APEC Summit in late November 2017 in Da Nang, President Trump directly emphasized that Indo – Pacific region would be "a place where sovereign and independent nations, with diverse cultures and many different dreams, can all prosper side-by-side, and thrive in freedom and in peace" (White House, 2017). Remarkably, one of the Free and Open Indo – Pacific Strategy's objectives is to preserve vitality of free and open air and maritime space in the region. The commencement speech by President Trump at the U.S. Naval Academy in 2018 highlighted the U.S. as a maritime nation. That is why the U.S., according to President Trump must always dominate that sea and oceans". Actually, the influence of sea power was first envisioned by the U.S. well-known naval historian, strategist and geopolitical theorist, Alfred Thayer Mahan. As stated by Mahan, control of seaborne can determine the winner and loser of wars. Since then, this ideology has been traditionally recognized by multiple U.S. Presidents. Nowadays, U.S. sea power, constituting several interrelated capabilities, including forward presence, deterrence, sea control, power projection, maritime security, humanitarian aid (CFR, 2019), bears not only military but also diplomatic significance for U.S. to maintain its superpower status.

At national state level, U.S. President is the leading, but not the sole, player in formatting and implementing the U.S. foreign policy. This process is largely joined by members of the Cabinet. Also, the U.S. Congress, together with other players, involves in formatting the U.S. foreign policy under "checks and balances" (Freegman, 1971: 35).

Though not a claimant in the East Sea, the U.S. Administration, including the Trump Administration, has long asserted its vital national interest in this waterway. U.S. Under Secretary of Defense for Acquisition and Sustainment Alan Shaffer believed that China's military build-up was threatening U.S. and allied interests in the Western Pacific and in the East Sea in particular (U.S. Department of State, 2020). Vice President Mike Pence emphasized that "Beijing's policies most harmful to America's interests and values, from China's debt diplomacy and military expansionism" (White House, 2019). In general, U.S. national interests in the East Sea can be felt in economic, military and strategic terms.

Economically, the United States' merchant shipping is less reliant on the East Sea, in comparison with China and Japan, with just over 14 percent (CSIS). However, as a super power of energy, the U.S. heavily depends on external supply of resources. Accounting for only 4.6 percent of the world's population, the U.S. consumes up to 25 percent of the world's oil. This fact, consequently, has led to "its strategic vulnerability" and constrained "its ability to pursue foreign policy and national security objectives" (CFR, 2006). Also, the oil shock in 2008 showed how energy security means to the U.S. giant economy indeed. While developing alternative resources of energy, including schist, oil and natural gas remain on the top of the list. In order to safeguard energy stability, the US is obliged to stretch out of its territory to compete with others in the fight for exploring foreign potential reserves, including the East Sea's reserves. China's ambiguous "nine dash line" claim in the East Sea comprises nearly 80% of the whole sea. At a joint conference in Manila in February 2019, U.S. Secretary of State Mike Pompeo pointed out that "China's island building and military activities in the East Sea threaten (Philippine) sovereignty, security and therefore economic livelihood, as well as that of the United States" (Cabato & Mahtani, 2019). Moreover, the U.S. has interests in securing unimpeded lawful trade through the East Sea, as repeated by many high-level officials.

Militarily and strategically, the East Sea is an extremely important sea lane to the U.S.'s defense networks and security links. On the one hand, the U.S. needs to bring security assurance to its regional allies, namely the Philippines and Taiwan, both of which make territorial claims in the East Sea. On the other hand, against the backdrop of China's rapid military build-up, the US needs to deter

China's desire from controlling the East Sea and protecting the rule-based order in the region. Every year, the U.S. conducts multiple bilateral and multilateral joint exercises, including the largest one in Asia, Cobra Gold. Then, any impediment to the free flow of both merchant and military shipping in the East Sea is a nuisance to the U.S. However, China's recent military build-up with an ultimate aim to replace the U.S. in the region has presented a credible threat to the U.S.'s national interests. According to James Fanell, a former Navy intelligence officer, "China will have about 550 warships by 2030 — nearly double the size of today's U.S. Navy" (Dorell, 2018). Since 2013, China has increasingly engaged in unprecedented and massive dredging and artificial island reclamations in the East Sea, expanding 3,200 acres of new land. More dangerously, those outposts have been significantly installed with long-range sensor arrays, port facilities, runways, and reinforced bunkers for fuel and weapons. The U.S. Department of Defense's 2019 annual report acknowledged that China has not conducted any new activities of militarization since its placement of air defense and anti-ship missiles in the Spratly islands in 2018. However, it is not, indeed, a positive signal to give the U.S. in particular and the region in general a big relief. Gen. Joseph Dunford, the Chairman of the U.S. Joint Chiefs of Staff in the conversation with the Brooking Institution in May 2019 insisted that if China's military build-up had peaked, it was because China had remarkably achieved its immediately military goals. Now, China is believed to have enough military capability "to monitor rivals' air and sea movements" and its artificial outposts can be utilized "as a base for coast guard and maritime militia operations against those countries' fishermen and hydrocarbon exploitation" (Stashwick, 2019). Consequently, it has posed threats to not only the U.S. navy's operations but also a rule-based order in the region which has been terribly destroyed by China's illegitimate unilateral expansionism.

Acknowledging national interests in the East Sea, the Trump Administration has shown its consensus in boosting proactive engagement in the Indo – Pacific in general and the East Sea in particular. The National Security Strategy (NSS) 2017, the National Military Strategy 2018 (NMS 2018) and the National Defense Strategy (NDS 2018) (White House, 2017), the most U.S. important national strategies, stated why the U.S. should maintain and increase its active engagement in the region. NSS 2017 realized that China's efforts "to build and militarize outposts in the East Sea" endangered "the free flow of trade", threatened "the sovereignty of other nations", and undermined "regional stability". Then, the NSS 2017 reaffirmed the U.S's commitments to freedom of the seas and the peaceful resolution of territorial and maritime disputes in accordance with international law". Meanwhile, the National Military Strategy 2018 (Joint Chief of Staff, 2018) (NMS 2018) pointed out that the reemergence of great power such as China and Russia has posed the most difficult challenges to the U.S. The National Defense Strategy (NDS 2018) highlighted that "China is leveraging military modernization, influence operations, and predatory economics to coerce neighboring countries to reorder the Indo-Pacific region to their advantage" (U.S. Department of Defense). Members of the U.S. Cabinet and other defense officials have also reaffirmed the U.S.'s rights of freedom of navigation. At the East Asia Summit in Singapore in November 2018, Vice President Mike Pence, while criticizing China's militarization and territorial expansion in the East Sea as "illegal and dangerous", reaffirmed that the U.S. "will continue to fly and sail wherever international law allows and our national interests demand. Harassment will not deter us; it only strengthens our resolve" (White House, 2019). In November 2019, Secretary of Defense Mark Esper stated that the U.S. had conducted "more freedom of navigation operations in the past year or so than we have in the past 20-plus year" (CFR, 2019). US Navy Commander Reann Mommsen, a spokesperson for the US 7th Fleet affirmed that "The U.S. will fly, sail and operate wherever international law allows", adding that freedom of navigation operations "are not about any one country, nor are they about making political statements" (Browne & Lendon, 2019). Commander Clay Doss, a U.S. 7th Fleet spokesman shared the view that "U.S. Forces operate in the Indo-Pacific region on a daily basis, including in the East Sea. All operations are designed in accordance with international law and demonstrate that the United States will fly, sail and operate wherever international law allows. That is true in the East Sea as in other places around the globe" (Werner, 2019). Also, both the U.S. Department of Defense and Department of State have helped build maritime capacity for Southeast Asian countries through the Indo – Pacific Maritime Security Strategy in June 2019 and the Indo – Pacific Transparency Initiative in November 2019. Additionally, the Indo – Pacific Strategy, including the East Sea policy, has been receiving strong bipartisan support. This was vigorously demonstrated by its "passage of the State and Foreign Operations, and Related Programs appropriations bill as a part of the end-of-year spending package, which contained $2.5 billion to implement the Gardner-Markey Asia Reassurance Initiative Act (ARIA)" (U.S. Senate, 2019). Highlighting the passage, Senator Gardner believed that ARIA "will ensure the United States remains the pre-eminent Pacific power for generations to come", "will allow the U.S. government to speak with one voice to advance our nation's national security, economic interests, and values in the Indo-Pacific, a region critical to the success of our nation as the pre-eminent global superpower that respects human rights and the rule of law" (U.S. Senate, 2019). ARIA is to reaffirm both the Trump Administration and U.S. Congress's commitments to "freedom of navigation under the international law" and the "peaceful resolution of maritime and territorial disputes" (CRS, 2019). The 2019 National Defense Authorization Act (NDAA) perceived China as its strategic competitor while emphasizing China's intensive militarization and land reclamation in the East Sea.

At systematic/ international level, as a superpower, the US is strongly affected by regional and international factors in drafting its foreign policy. Any change in the balance of power in the disadvantageous vector for the U.S. or any nation-state' rise challenging the U.S. already set-world order will force the U.S. to recalculate its strategies to restore its power and influence. Also, as a superpower, the U.S. should express its responsibilities in paying respects for and legally abided by international law and ruled-based orders. The U.S.'s National Security Strategy (NSS 2017) and National Defense Strategy 2018 (NDS 2018) both labeled China (and Russia) as the U.S.'s leading strategic competitor in the region. In the Indo – Pacific, China's rise has been posing risks in various fields, including maritime security. Specifically, China's recent developments, including its unceasingly reclamations and militarization of artificial islands in the East Sea in order to legalize its "nine-dashed line" have been threatening regional rule-based order as well as the U.S.'s economic, military and strategic interests. Vice President Mike Pence at the 13th EAS in November 2018 stated that "our commitment to uphold the freedom of the seas and skies, where we stand shoulder to shoulder with you for freedom of navigation" (U.S. Embassy in the Republic of Korea, 2018). It is added that "China's militarization and territorial expansion in the East Sea is illegal and dangerous, threatens the sovereignty of many nations and endangers the prosperity of the world" (U.S. Embassy in the Republic of Korea, 2018). Commander, US. Pacific Fleet, Admiral Scott H. Swift once emphasized that "China is challenging that principle (the principle of unfettered access to the shared global spaces for all nations) across all elements of national power characterized by the acronym DIME: Diplomatic, Information, Military and Economic" (U.S. Navy, 2017), adding that "freedom of navigation operations serve to reassert the inviolability of shared spaces and reaffirms America's commitment to upholding the rules-based international system" (U.S. Navy, 2017).

4. US's policy implementation in the East Sea

US's policy implementation in the East Sea can be seen mostly at national level and systematic levels under the leadership of President Trump.

At national level, with the slogan of "peace through strength", since the very beginning, the Trump Administration has focused on military build-up and rotation. In order to secure peace, stability and prosperity in the region, Chief of Naval Operations Admiral John Richardson put forth a vision in early 2017 for the U.S.'s Future Navy, in which "the nation needs a more powerful Navy, on the order of 350 ships, that includes a combination of manned and unmanned systems" (Maritime Issues, 2017). At the Shangri-La Dialogue, Secretary of Defense James Mattis declared in November 2017 that "currently 60% of all US Navy ships, 55% of Army forces and about two-thirds of Fleet Marine forces are assigned to the US Pacific Command area of responsibility. Soon, 60% of our overseas tactical-aviation assets will be assigned to this theatre." (Maritime Issues, 2017). In adaption to changing circumstances in the Indo – Pacific region, Secretary of Defense James Mattis officially "rename the US Pacific Command to US Indo - Pacific Command" in May 2018, which has about 375,000 civilian and military personnel, "more of the world than any of the five over geographic combatant commands and shares a border with each of its counterparts" (PACOM). Also, in August 2019, Secretary of Defense Mark Esper said that the U.S would invest in more bases in the region, "adding to its China containment activities in the region" (Jeong-ho & Ng, 2019). Admiral Philip S. Davidson, Commander of the U.S. Indo-Pacific Command in a hearing in February 2020 also revealed possibility of revisiting some of the places that the U.S. has operated and rotated forces. At systematic level, the U.S. continues to actively engage in key multilateral mechanism such as the ASEAN Regional Forum, the ASEAN Defense Ministers Meeting-Plus, and the East Asia Summit. Remarkably, the U.S. has utilized regional mechanisms to blame China's unlawful activities. Acting U.S. Defense Secretary Patrick Shanahan noted at the Shangri-La Dialogue in 2019 that the U.S. "will continue to support the freedom of navigation, free and open Indo – Pacific" while indirectly criticizing China for "toolkit of coercion" in the East Sea (CNN, 2019). Meanwhile, U.S. Envoy Robert O'Brien at the ASEAN – U.S. Summit in Thailand in

November 2019 emphasized that China "has used intimidation to try to stop ASEAN nations from exploiting the off-shore resources, blocking access to 2.5 trillion dollars of oil and gas reserve alone" (Bankok Post, 2019). During the Munich Conference in February 2020, Secretary of Defense Mark Esper noted China's seizing and militarizing artificial islands in the East Sea which would "alter the landscape of power and reshape the world in their favor ….and often at the expense of others" (U.S. Department of Defense, 2020). Not only criticizing China's behavior, the U.S. showed support for exploiting legal measures in addressing territorial disputes in the East Sea, including early conclusion of Code of Conduct between China and ASEAN as well as its respect for the Tribunal ruling. On July 13, 2020, Secretary of State Mike Pompeo announced that China's expansive maritime claims in the East Sea were "completely unlawful". Though the U.S. has affirmed that it has not changed its neutrality policy on competing claims to legitimate land features in the East Sea, its new position has been already the strongest and most explicit support of the 2016 ruling. For the very first time, the U.S. has involved itself in the legal battle of diplomatic note exchanges between China and other claimants. Particularly, in June 2020, the U.S. Ambassador to the U.S. officially sent out a letter to the Secretary-General of the United Nation which reiterated its objections to China's maritime claims in the East Sea.While consolidating and deepening a network of allies and partners, the U.S. seeks to develop new partnerships with "pivotal players across the region, such as Indonesia, Malaysia, and Vietnam" to "address common challenges, to enhance shared capabilities, to increase defense investment where appropriate, to improve interoperability, to streamline information sharing, and to build networks of capable and like-minded partners" (U.S. Department of Defense, 2018). In the relationship with its oldest ally in the region, the Philippines, joint military activities was planned to increase in 2019 from 262 to 281 (Heritage Foundation, 2019). Remarkably, the Trump Administration publicly declared its security protection of the Philippines for the very first time. During his visit to the Philippines in February 2019, Secretary of State Mike Pompeo affirmed that "As the East Sea is part of the Pacific, any armed attack on Philippine forces, aircraft or public vessels in the East Sea will trigger mutual defense obligations under Article 4 of our mutual defense treaty" (Cabato & Mahtani, 2019). In the context that President Duterte planned to terminate the Visiting Forces Agreement, Admiral Philip S. Davidson, Commander of the U.S. Indo-Pacific Command expressed hope that the U.S. Department of State would be able to negotiate a solution that would secure the Visiting Forces Agreement while reaffirming that the 1951 Mutual Defense Treaty would be applied for excessive territorial claims in the East Sea. In the meantime, the Trump Administration has boosted its regional maritime capacity building efforts through the Indo – Pacific Maritime Security Initiative (U.S. Department of Defense, 2019) with increasing and extending funding till 2024 and the Indo – Pacific Transparency Initiative (U.S. Department of State, 2018). According to the Trump Administration, in the year of 2018, the US sold US$ 9.42 billion worth of arms and provided more than US$500 million in security assistance (more than double the previous year) to regional states. Moreover, for the first time ever, the U.S. conducted joint military exercises with ASEAN in September, 2019 to enhance capacities of ASEAN nations' naval forces in the fight against naval and disaster threats (Harmer, 2019). To enhance

U.S. presence in the region, ARIA is set to authorize $1.5 billion annually for the period from 2019 to 2023, especially to address security concerns such as China's aggressive actions in the East Sea (CogitAsia, 2019).

The. U.S. has been straight-forward in communicating with China on the East Sea. After a high-level talk with Chinese counterpart in November 2018, Secretary of State Mike Pompeo highlighted the U.S. concerns about China's activities in the East Sea and "press China to live up its past commitments" of "non-confrontation" (CNBC, 2018) in the region. U.S. National Security Adviser John Bolton in 2019 strongly criticized China's actions in the East Sea as "completely unacceptable" and would "continue to pursue actions to prevent Beijing from turning the area into a new Chinese province" (ABC News, 2020). It is noteworthy to mention that the U.S. expressed its concern not only over China's illegal reclamation and militarization but also its interference with oil and gas activities in the East Sea, including Vietnam's long-standing exploration and production activities. In July 2019, Department Spokesperson Morgan Ortagus stated that "China's repeated provocative actions aimed at the offshore oil and gas development of other claimant states threaten regional energy security and undermine the free and open Indo-Pacific energy market" (U.S. Department of State, 2019). Then, the U.S. strongly believed that China should stop bullying its neighbors and refrain from provocative and destabilizing activities. Recently, upon China's sinking of a Vietnamese vessel in the vicinity of the Paracel islands in the East Sea, the U.S. has been among the first condemning China's asserting "unlawful maritime claims and disadvantage its Southeast Asian neighbors" in this water (U.S. Department of State, 2020). The U.S. Department of Defense highlighted that China's behavior "stands in contrast to the U.S.'s vision of a free and open Indo - Pacific region, in which all nations, large and small, are secure in their sovereignty, free from coercion, and able to pursue economic growth consistent with accepted international rules and norms" (U.S. Department of Defense, 2020). Instead of destabilizing the region, the U.S. called for focusing on taming the corona pandemic. Also, the Trump Administration has sought new ways to reduce risks of naval encounters in disputed waters, including the East Sea. Admiral John Richardson, Chief of US Naval Operations, at a meeting of the Atlantic Council in 2019, called for firmer rules governing encounters not only between navies but also coastguards and maritime militias, "so-called second and third sea forces that Beijing has used to advance its sovereignty claims" (Zhou, 2019). Previously in 2014, China and the U.S. agreed to the Code for Unplanned Encounters at Sea (CUES), a non-legally binding accord that was limited to reducing escalation of tensions and chances of clashes between naval vessels and military aircrafts only. So far, the two sides and other foreign Navies have not kicked off any new round of negotiation. Speaking with his counterpart in March 2020, Secretary of Defense Mark Esper expressed his concern over China's lasing U.S. Navy P-8A Poseidon maritime patrol aircraft west of Guam which violates the Code for Unplanned Encounters at Sea (CUES). Then, he raised the need for enhancing bilateral communication mechanisms to resolve escalation of crisis.

Throughout its history, the U.S. has conducted FONOPS, as recognized in the UNCLOS 1982, to stop coastal states with illegally excessive maritime claims from infringing its rights, freedom, and lawful uses of the sea and explorations of common goods in the sea. Formally established in

1979, the FONOP is implemented by both the U.S. Department of State and Department of Defense in order to preserve the U.S. vital national interests in the seas. While the former leads diplomatic efforts to protest excessive maritime claims, the later carries out operational assertions against excessive maritime claims. The Trump Administration conducted the first FONOP very early, just four months after taking office. The first FONOP within twelve nautical miles around Scarborough Shoal since China seized it in 2012 was conducted in January 2018. U.S. Navy's sailing within 12 nautical miles of features claimed or occupied by China hit the highest record of nine times in 2019 since China started to conduct land reclamations in 2014. Previously, such FONOPs were carried five in 2018, six in 2017 under the Trump Administration. Also, the US. is no longer a single operator conducting FONOPs in the waterway. There have been other allies and partners joining with the U.S. Navy, namely warships of the Royal Australian Navy (RAN) in November 2019.

5. Conclusion

So far, there has not been any U.S. formal paper on the East Sea. However, the extent that President Trump and other senior officials have attached importance to the issue, at both individual and national level, in order to constrain and deter China's aggressiveness in the East Sea at regional level, has proved that the U.S. has indeed had a policy in this waterway. By applying three levels of analysis, it can be clearly seen how internal and external factors have had impact on the Trump's East Sea policy. On the one hand, the Trump Administration will continue to push "America's First" to serve for the upcoming Presidential Election, especially against the background of the novel corona pandemic's huge impacts on the country. On the other hand, though being a non-claimant in the East Sea and a non-UNCLOS member, the Trump Administration will maintain its decoupling with China in various areas, including the East Sea, in order to secure a "free and open Indo – Pacific". Hence, it is critical to closely observe US's move regarding the East Sea, then proactively and timely coordinating with other related countries, especially in the year Viet Nam undertaking ASEAN Chairmanship.

Reference

ABC news (2020), "Recent developments surrounding East Sea", https://abcnews.go.com/International/wireStory/recent-developments-surrounding-south-china-sea-61602195, retrieved on March 2019.

Academia, *Chapter three on Levels of Analysis and Foreign Policy,* https://www.academia.edu/28941706/Chapter_Three_on_Levels_of_Analysis_and_Foreign_Policy.

Bangkok Post (2019), "US Envoy decries Chinese intimidation in East Sea", https://www.bangkokpost.com/thailand/general/1786889/us-envoy-decries-chinese-intimidation-in-south-china-sea, retrieved on November 2019.

Ben Werner (2019), "Two U.S. Guided-Missile Destroyers Conduct FONOP Past Mischief Reef in East China Sea", *USNI,* https://news.usni.org/2019/02/11/41017, retrieved on 11 February 2019.

Business Insider (2019), "US flies B-52 bombers over East Sea for the 2nd time in 10 days",https://www.businessinsider.com/us-flies-b-52-bombers-over-south-china-sea-for-the-2nd-time-in-10-days-2019-3, retrieved on March 2019.

CFR (2019), "Sea power, US Navy and Foreign Policy", https://www.cfr.org/backgrounder/sea-power-us-navy-and-foreign-policy, retrieved on August 2019.

CFR (2019), "US – China Strategic Competition in South and East China Seas: Background and Issues for Congress", https://fas.org/sgp/crs/row/R42784.pdf, retrieved on 26 November 2019.

CFR (2006), "US Energy Dependence Undercutting U.S. National Security, Council Task Force Warns", https://www.cfr.org/news-releases/us-energy-dependence-undercutting-us-national-security-council-task-force-warns, retrieved on October 2006.

China Power CSIS, "How much trade transits the East Sea?",https://chinapower.csis.org/much-trade-transits-south-china-sea/

CNBC (2018), "US presses China to halt militarization of East Sea", https://www.cnbc.com/2018/11/10/us-presses-china-to-halt-militarization-of-south-china-sea.html, retrieved on November 2018.

CNN (2019), "Singapore Shangrila defense conference", https://edition.cnn.com/2019/05/31/politics/singapore-shangrila-defense-conference/index.html, retrieved on June 2019.

CogitAsia (2019), "The Asia Reassurance Initiative Act overlooking smaller players in Southeast Asia", https://www.cogitasia.com/the-asia-reassurance-initiative-act-overlooking-smaller-players-in-southeast-asia/, retrieved on April 25, 2019.

CRS (2019), "ARIA", https://fas.org/sgp/crs/row/IF11148.pdf, retrieved on April 2019.

Heritage Foundation (2019), "Assessing the Global Operating Environment", https://www.heritage.org/military-strength/assessing-the-global-operating-environment/asia, retrieved on August 2019.

J. David Singer (1961), "The Level-of-Analysis Problem in International Relations", *World Politics,* Vol. 14, No. 1, pp. 77-92.

John Ikenberry (2014), *"American Foreign Policy"*, Oxford University, 7th edition, page 3.

Joint Chief of Staff (2018), "National Military Strategy", https://www.jcs.mil/Portals/36/Documents/Publications/UNCLASS_2018_National_Military_Strategy_Description.pdf, retrieved on June 2018.

Leonard Freegman (1971), Roger A. Riske, *"Power and Politics in America",* 5th edition, p. 35.

Maritime Issues (2017), "The US naval build-up and Asia under the Trump Administration, http://www.maritimeissues.com/us-new-administration/the-us-naval-buildup-and-asia-under-the-trump-administration.html, retrieved on July 2017.

Jerry Harmer (2019), "US promotes Free and Open Indo-Pacific at naval exercise", https://www.navytimes.com/news/your-navy/2019/09/02/us-promotes-free-and-open-indo-pacific-at-naval-exercise/, retrieved on September 2nd, 2019.

Oren Dorell (2018), "How China's military expansion threatens U.S. interests", *USA Today,* https://www.usatoday.com/story/news/world/2018/05/18/how-chinas-military-expansion-threatens-u-s-interests/621385002/, retrieved on May 2018.

PACOM, US PACOM Area of Responsibility, https://www.pacom.mil/About-USINDOPACOM/USPACOM-Area-of-Responsibility/

Regine Cabato,Shibani Mahtani (2019), "Pompeo promises intervention if Philippines is attacked in East Sea amid rising Chinese militarization", *Washington Post,* https://www.washingtonpost.com/world/pompeo-promises-intervention-if-philippines-is-attacked-in-south-china-sea-amid-rising-chinese-militarization/2019/02/28/5288768a-3b53-11e9-b10b-f05a22e75865_story.html, retrieved on February 2019.

Rourke, J.T. & Boyer M.A. (2010), *"International politics on the world stage",* McGraw Hill Higher Education, 8th edition, Chapter 3.

Ryan Browne, Brad Lendon (2019), "US sails warship near contested islands in the East China Sea amid tension with China", *CNN,* https://edition.cnn.com/2019/08/28/politics/us-navy-south-china-sea/index.html, retrieved on 29 August 2019.

Lee Jeong-ho, Teddy Ng, (2019), "New US military bases in Asia-Pacific 'likely to be temporary' for troop flexibility", https://www.scmp.com/news/china/diplomacy/article/3025125/new-us-military-bases-asia-pacific-likely-be-temporary-troop, retrieved on August 2019.

Laura Zhou (2019), "East Sea Trump military adviser calls firmer rules to stop near misses", https://www.scmp.com/news/china/diplomacy/article/2185308/south-china-sea-trump-military-adviser-calls-firmer-rules-stop, retrieved on February 2019.

Steven Stashwick (2019), "China's East Sea Militarization Has Peaked", *Foreign Policy,* https://foreignpolicy.com/2019/08/19/chinas-south-china-sea-militarization-has-peaked/, retrieved on August 2019.

The Atlantic (2018), "Read President Trump's U.S. Naval Academy Commencement Address, https://www.theatlantic.com/politics/archive/2018/05/read-president-trumps-us-naval-academy-commencement-address/561206/, retrieved on May 2018.

US State Department (2018), "Indo-Pacific Transparency Initiative", https://www.state.gov/indo-pacific-transparency-initiative/, retrieved on November 3rd 2019.

US Department of State (2019), "Press Briefing", https://www.state.gov/briefings/department-press-briefing-july-11-2019/, retrieved on July 2019.

U.S. Department of State (2019), "Press Statement", https://www.state.gov/chinese-coercion-on-oil-and-gas-activity-in-the-south-china-sea/, retrieved on July 2019.

US Department of State (2020), "Press Statement", https://www.state.gov/prcs-reported-sinking-of-a-vietnamese-fishing-vessel-in-the-south-china-sea/, retrieved on April 2020.

U.S. Department of Defense (2019), "China a rising threat to national security, says DOD leaders", https://www.defense.gov/Explore/News/Article/Article/1784442/china-a-rising-threat-to-national-security-say-dod-leaders/, retrieved on March 2019.

US. Department of Defense (2020), "Immediate Release", https://www.defense.gov/Newsroom/Releases/Release/Article/2143925/china-coast-guard-sinking-of-a-vietnam-fishing-vessel/, retrieved on April 2020.

US Department of Defense (2019), Indo – Pacific Strategy Report, https://media.defense.gov/2019/Jul/01/2002152311/-1/-1/1/DEPARTMENT-OF-DEFENSE-INDO-PACIFIC-STRATEGY-REPORT-2019.PDF, retrieved on June 1st 2019.

U.S. Department of Defense (2018), "National Defense Strategy", https://dod.defense.gov/Portals/1/Documents/pubs/2018-National-Defense-Strategy-Summary.pdf, retrieved on January 2018.

US. Department of Defense (2020), "Remarks by Secretary of Defense at the Munich Conference", https://www.defense.gov/Newsroom/Speeches/Speech/Article/2085577/remarks-by-secretary-of-defense-mark-t-esper-at-the-munich-security-conference/, , retrieved on February 2020.

US Department of Defense (2018), "Remarks by Secretary Mattis at Plenary Session of the 2018 Shangri-La Dialogue", https://www.defense.gov/Newsroom/Transcripts/Transcript/Article/1538599/remarks-by-secretary-mattis-at-plenary-session-of-the-2018-shangri-la-dialogue/, retrieved on June 2nd 2018.

US Embassy in the Republic of Korea (2018), "Remarks by Vice President Pence at the 6th US-ASEAN Summit", https://kr.usembassy.gov/111418-remarks-by-vice-president-pence-at-the-6th-u-s-asean-summit/, retrieved on November 2018.

US Navy (2017), "Future of Asia", https://www.cpf.navy.mil/leaders/scott-swift/speeches/2017/09/future-of-asia.pdf, retrieved on September 2017.

U.S. Senate (2019), "Press Release", https://www.gardner.senate.gov/newsroom/press-releases/gardner-markey-applaud-25b-in-government-funding-bill-to-implement-indo-pacific-policy, retrieved on December 2019.

White House (2018), "Prepared Remarks for Vice President Pence at the East Asia Summit Plenary Session", https://www.whitehouse.gov/briefings-statements/prepared-remarks-vice-president-pence-east-asia-summit-plenary-session/, retrieved on 15 November 2018.

White House (2017), "National Security Strategy", https://www.whitehouse.gov/wp-content/uploads/2017/12/NSS-Final-12-18-2017-0905-2.pdf, retrieved onDecember 2017.

White House (2017), "Remarks by President Trump at APEC CEO Summit in Da Nang, Viet Nam", https://www.whitehouse.gov/briefings-statements/remarks-president-trump-apec-ceo-summit-da-nang-vietnam/, retrieved on November 2017.

White House (2019), "Remarks by Vice President Pence", https://www.whitehouse.gov/briefings-statements/remarks-vice-president-pence-frederic-v-malek-memorial-lecture/, retrieved on October 2019.

CÁC CẤP ĐỘ PHÂN TÍCH TỪ LÝ THUYẾT ĐẾN THỰC TIỄN - TRƯỜNG HỢP NGHIÊN CỨU ĐIỂN HÌNH: CHÍNH SÁCH BIỂN ĐÔNG CỦA MỸ DƯỚI CHÍNH QUYỀN TRUMP

Phạm Minh Thu

Học viện Ngoại giao Việt Nam, Vụ Châu Mỹ, Bộ Ngoại giao
69 Chùa Láng, Đống Đa, Hà Nội, Việt Nam

Tóm tắt: Ba cấp độ phân tích (cá nhân, quốc gia và hệ thống quốc tế) được công nhận rộng rãi trong phân tích chính sách đối ngoại có thể được vận dụng để lý giải liệu Chính quyền Trump có thực sự có chính sách Biển Đông hay không. Ở cấp độ cá nhân, Tổng thống Trump lần đầu tiên công bố "Chiến lược Ấn Độ - Thái Bình Dương tự do và mở" của Mỹ vào tháng 11/2017, theo đó Biển Đông được coi là nhân tố quan trọng của trụ cột an ninh nhằm duy trì vùng trời và không gian hàng hải tự do và mở ở khu vực. Ở cấp độ quốc gia, Biển Đông tiếp tục được nhắc lại trong các Chiến lược quốc gia của Mỹ gồm Chiến lược An ninh Quốc gia, Chiến lược Quốc phòng Quốc gia và Chiến lược Quân sự Quốc gia. Bên cạnh đó, các thành viên trong nội các Mỹ và các nghị sĩ Quốc hội Mỹ cũng thể hiện sự ủng hộ đối với Chiến lược Ấn Độ - Thái Bình Dương nói chung và chính sách Biển Đông của Mỹ nói riêng. Để đương đầu với sự hung hăng trên biển của Trung Quốc ở cấp độ hệ thống, Chính quyền Trump đã không ngừng triển khai sức mạnh đáng kể, bao gồm việc tăng cường hiện diện quân sự, hỗ trợ xây dựng năng lực hàng hải cho khu vực cũng như triển khai các hoạt động tự do hàng hải (FONOP) thường xuyên. Trong bối cảnh tranh chấp lãnh thổ tại Biển Đông tiếp tục diễn biến phức tạp và khó lường, việc đánh giá kỹ lưỡng chính sách Biển Đông của Chính quyền Trump sẽ có ý nghĩa đối với các nước có yêu sách ở vùng biển này, trong đó có Việt Nam.

Từ khóa: Các cấp độ phân tích, Trump, chính sách, Biển Đông.

INSTRUCTIONS TO AUTHORS

1. As a natural development from its predecessor *VNU Journal of Science: Foreign Studies*, VNU Journal of Foreign Studies continues to be an official, independent peer-reviewed publication of the University of Languages and International Studies (ULIS) under Vietnam National University (VNU). *VNU Journal of Foreign Studies* publishes bimonthly in 2 Vietnamese editions and 4 English editions p.a. with research articles in *linguistics*, *language/foreign language teaching*, *international studies*, and *related social sciences and humanities*. Only papers which have not been previously or elsewhere published can be accepted.

2. The manuscript should cite AT LEAST 01 article published in the VNU Journal of Foreign Studies.

3. All manuscripts undergo a double-blind peer review process, so authors need to avoid unnecessary disclosure of their identities in content.

4. Manuscripts should be typed in Vietnamese or English (at least 10 pages for research article and 5 pages for discussion article), using A4 size, left margin 2.5cm, right margin 2.5cm, top margin 3.6cm, bottom margin 4.3cm, Times New Roman font at 11 with Single spacing.

5. All manuscripts must contain abstract (200 words maximum) and at most five (5) keywords in both Vietnamese and English. Article title should be bold and capitalized.

6. Manuscripts should be prepared according to the following order: Article title, Author(s), Workplace(s), Abstract, Keywords, Main text, Acknowledgements (if any), References, Appendix (if any). The corresponding author should be marked, with his/her postal address, e-mail, telephone provided on the first page of the manuscript.

7. Sections and subsections must be clearly defined and numbered. Main sections should be numbered 1, 2, 3, etc. Subsections must be numbered 1.1, 1.2, 1.3, 2.1, 2.2, 2.3, etc. (then 1.1.1, 1.1.2, 1.1.3, etc.). The abstract is not included in section numbering.

8. Illustrations must be clear and numbered according to the sequence in the article. The sources of the illustrations must be clearly indicated. Ensure that each illustration has a title. The titles of the illustrations must be provided below the illustrations.

E.g.:

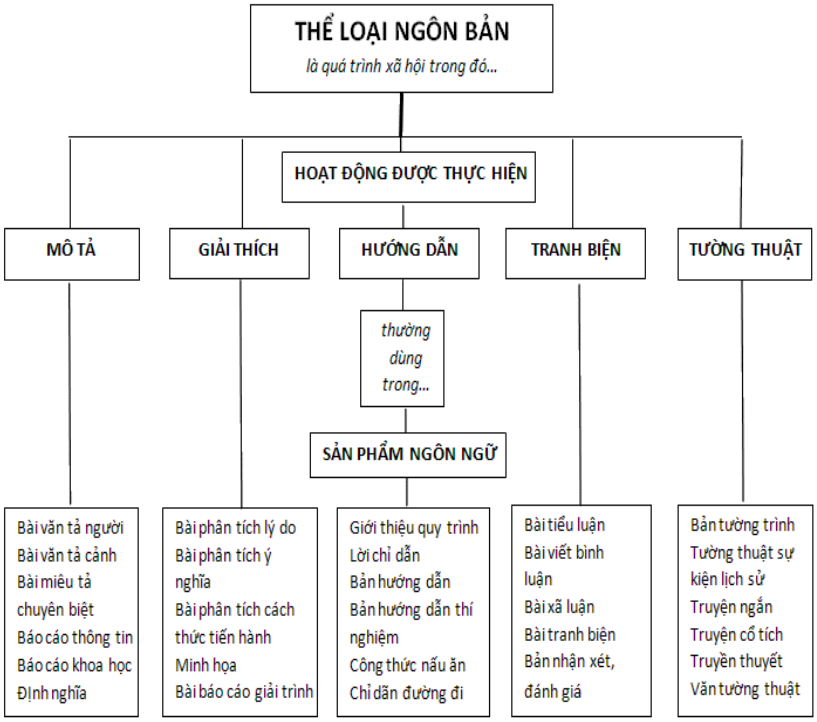


Figure 1. Classifications of text types (adapted from Knapp & Watkins, 1994)

8. Tables should be submitted as editable text and not as images. Tables must be consecutively numbered in accordance with their appearance in the article. The titles of the tables must be provided above the tables. Please avoid using vertical rules in tables.

E.g.:

Table 1. Effectiveness of applying exploratory learning model in teaching Education subject

STT	**Giá trị đạt được**	$\bar{X}$	***SD***
1	Giá trị thực tiễn của những giờ học môn Giáo dục học qua các hình thức học tập khám phá	1,8070	,65008
2	**Đáp** ứng yêu cầu thực tiễn phát triển kỹ năng nghề	1,8947	,68315
3	Góp phần hình thành năng lực hành **động** cho sinh viên	1,7807	,64847
4	Nâng cao hứng thú học môn Giáo dục học cho sinh viên	1,9386	,81235

9. References

9.1. In-text citation:

- If the author(s) is (are) Vietnamese:

+ Cite full name of the author(s) according to the order in Vietnamese (surname, middle name, given name) if the reference cited is Vietnamese.

E.g.: Ở Việt Nam những năm gần đây, nghiên cứu về chữ Hán phải nhắc đến Chữ Hán: chữ và *nghĩa* (Phạm Ngọc Hàm, 2012)

+ Cite surname of the author(s) if the reference cited is English.

E.g.: Several reasons are believed to hold accountable for this trend, namely the low salary, the poor career prospect as well as the pressure from huge workload and students' parents (Nguyen, 2019).

- If the author(s) is (are) not Vietnamese, only their surnames are mentioned in the article. In case it is difficult to identify which is the surname and which is the given name of the author, it is advisable that his/her full name be provided.

- For direct citation, provide the page(s) of the reference cited (Use "p./pp." in English article and "tr." in Vietnamese article).

E.g.:

According to Jones (1998), "students often had difficulty using APA style, especially when it was their first time" (p. 199).

She stated, "Students often had difficulty using APA style" (Jones, 1998, p. 199), but she did not offer an explanation as to why.

Ví dụ, đối với nghệ thuật truyền thống, một mặt bảo tồn nguyên trạng những gì còn lại, đồng thời mặt khác có thể phát triển, thay đổi ở những mức độ sáng tạo khác nhau (Bùi Quang Thắng và cộng sự, 2012, tr. 51-55).

- If the article has two authors, cite their surname names and use ***and*** between them (in English articles*)*; cite their full names and use ***và*** between them (in Vietnamese articles*)*.

- If the article has more than two authors, use their surname names and ***et al.*** (in English articles); use their full names and ***và cộng sự*** (in Vietnamese articles*)*.

9.2. Reference list:

- Please ensure that every reference cited in the text is also present in the reference list. References should be clarified in the main text. References are provided in their original (published) language, with each language arranged in a separate section in alphabetical order, with Vietnamese given first and foremost, e.g. Vietnamese, Arabic, Chinese, English, French, German, Japanese, Korean, Russian and Thai.

- References are provided in their original (published) language, not in phonetic style.

E.g.:

范明进主编（2018）．汉字文化圈：汉语教学与研究．河内国家大学出版社, 25-30.

- References translated into Vietnamese are included in Vietnamese reference list.

- If the author is Vietnamese and the text is written in a foreign language, the reference of the text is included in foreign reference list (even the text is published in Vietnam).

Eg.:

Tran, T. C., & Do, T. T. H. (2015). Patterns of Code-Mixing of English in Hoa Hoc Tro Magazine in Vietnam. *VNU Journal of Science: Foreign Studies, 31*(4), 11-24.

- For Vietnamese references and Vietnamese authors, cite full names of the authors as follows:

E.g.:

Nguyễn Thiện Giáp (2008). *Giáo trình Ngôn ngữ học*. Hà Nội: Nxb Đại học Quốc gia Hà Nội.

Lâm Quang Đông (2015). Quá trình tư duy ở người lớn học ngoại ngữ: một trường hợp điển cứu. Kỷ yếu Hội thảo quốc tế *Đổi mới việc dạy-học & nghiên cứu Hàn Quốc học tại Việt Nam.* Hà Nội, ngày 10 tháng 4 năm 2015, tr. 172-191.

- For English references, conform to APA standard.

- Citations in the text and references should conform to referencing style used by the American Psychological Association, namely:

+ Books:

Saussure, Ferdinand de (1959). *Course in General Linguistics.* Edited by Charles Bally and Albert Sechehaye in collaboration with Albert Reidlinger, translated by Wade Baskin. New York: Philosophical Library.

Nguyễn Thiện Giáp (2008). *Giáo trình Ngôn ngữ học*. Hà Nội: Nxb Đại học Quốc gia Hà Nội

+ Book chapters:

Levi-Strauss, C. (1971). Totem and caste. In F. E. Katz (Ed.), *Contemporary sociological theory* (pp. 82- 89). Random House.

Phan Huy Đường, (2007), "Chương 3 – Phát triển kinh tế đối ngoại Việt Nam qua các giai đoạn lịch sử hiện đại", *Kinh tế đối ngoại Việt nam*. Hà Nội: Nxb Đại học Quốc gia Hà Nội, tr. 98-178.

+ Journal articles or conference papers:

Alptekin, C. (2002). Towards intercultural communicative competence in ELT. *ELT Journal*, *56*(1), 57-64.

Krashen, S. (1979). 'The Monitor Model for second language acquisition'. In R. Gingras (ed.) *Second Language Acquisition and Foreign Language Teaching*. CAL.

Kirkwood, M., & Christie, D. (2006). The role of teacher research in continuing professional development. *British Journal of Educational Studies*, *54*(4), 429-448.

Nguyễn Hữu Minh (2015). Gia đình Việt Nam sau 30 năm đổi mới. *Khoa học xã hội Việt Nam, 11*(96), 51-59.

Lâm Ngọc Như Trúc (2008). Công nghiệp hóa và sự biến đổi của gia đình Việt Nam. Kỷ yếu Hội thảo quốc tế *Vi yếu Hội t* lần III. Hà Nội, ngày 5 tháng 12 năm 2008, tr.163-187.

+ Online materials:

Baker, Mark C. (1989). Object Sharing and Projection in Serial Verb Constructions. *Linguistic Inquiry, 20*(4) (Autumn, 1989), 513-553, Cambridge: The MIT Press. Available through <http:// www.jstor.org/stable/ 4178644>, Accessed 14/03/2012 06:49.

An Ngọc (2013). Nghề biên-phiên dịch thiếu chuẩn mực nghề nghiệp, Truy cập lúc 10:30 ngày 29/11/2015 tại http://www.vietnamplus.vn/nghe-bienphien-dich-thieu-chuan-muc-nghe-nghiep/234990.vnp.

+ Videos:

Staveley-Taylor, H. (Director). (2006). *Introduction to designing experiments* [Film; DVD]. Uniview Worldwide; Cambridge Educational.

Hà Văn Tân (2018, 8 tháng 9). *Nghiên cứu khoa học là gì?* [Video]. Youtube. https://www.youtube.com/watch?v=nLkj95pgl28

+ Unpublished doctoral and master thesis:

Gerena, C. (2015). *Positive Thinking in Dance: The Benefits of Positive Self-Talk Practice in Conjunction with Somatic Exercises for Collegiate Dancers*, Master's thesis, University of California Irvine.

Nguyễn Hoàng Thanh (2011). *Nghiên cứu mức sẵn sàng chi trả cho cải thiện điều kiện vệ sinh môi trường tại huyện Kim Bảng, Hà Nam năm 2010,* Luận văn Thạc sĩ y tế công cộng, Trường Đại học Y Hà Nội.

10. Manuscripts should be sent to the editorial board at the email address below and will not be returned to the author(s) in case of rejection. The author(s) must bear any legal responsibility for the article and references.

***VNU Journal of Foreign Studies*, 3rd Floor, Building A1,**

VNU University of Languages and International Studies, #2 Pham Van Dong Street, Cau Giay District, Hanoi, Vietnam

Website: https://js.vnu.edu.vn/FS/

Email: tapchincnn@gmail.com / tapchincnn@vnu.edu.vn

ĐT.: (84-24) 62532956

STRUCTURE OF A MANUSCRIPT

TITLE

(In English and Vietnamese, capitalized, Constantia font at 16,
single line spacing, centering)

Author's name (Times New Roman font at 13.5)[1*]

Name of agency, university (Times New Roman font at 10,5, in italics)

Address of agency, university (Times New Roman font at 10,5, in italics)

Abstract: Abstract must be provided in both English and Vietnamese, no more than 200 words, Times New Roman font at 10pts.

Key words: No more than 5 words, Times New Roman font at 10 pts.

The main content of the manuscript contain the following sections (Times New Roman font at 11 pts, single line spacing).

1. Statement of problem

2. Theoretical background

2.1 …

2.2 ….

3. Research methodology

3.1 …

3.2 …

4. Findings

4.1 ….

4.2 ….

5. Discussion of findings

5.1 …

5.2 …

6. Conclusion

References

Appendix

---

[*] Tel.: Telephone number of the corresponding author

Email: (Email address of the corresponding author)