An Intra-Task Washback Investigation of a Speaking Test

Todd Armstrong

In the field of education, previous washback studies were conducted to investigate the effect tests had on educational systems. One type of study are those relating to traditional, multiple choice, large-scale tests, which are perceived to have had mainly negative influences on the quality of teaching and learning (Madaus & Kellaghan, 1992; Shepard, 1990). The second type involve studies in which a specific test or examination has been modified and improved upon in order to exert a positive influence on teaching and learning. This paper falls into the latter category.

This paper examines an intra-task comparison of two speaking tests. In an intra-task comparison, aspects of a particular task, such as preparation time to accomplish the task, are systematically manipulated. The research question was: Does having increased preparation time result in scripted dialogues between test takers? Data from seven pairs of students performing five-minute speaking examinations was gathered and analyzed. It will be shown that the more preparation time the test takers were given, the more scripted the conversations became. The pedagogical consequence of such time preparation goes to the very heart of communicative competence.

There are two terms that need to be defined; communicative competence, and washback. These terms will be defined in the next section.

Literature Review

One of the early definitions of communicative competence was proposed by Sauvignon (1972) who used it to describe the ability of classroom language learners to interact with other speakers, to make meaning, as distinct from their ability to recite dialogues or perform on discrete-point tests of grammatical knowledge. In other words, communicative competence is the ability to communicate in spoken or written form in the

language being learned. In many language classes around the world, language study is not viewed as a communication tool but as a body of knowledge, to be analyzed and studied in the learner's first language. From there, learners are to translate between their first language and the foreign language for the sole purpose of a written exam. These exams consist of discrete point type questions. It can be argued that without comprehensible input very little foreign language acquisition can occur. However, in many foreign language learning classrooms there is an overemphasis of grammar-translation in the learners' native language with little focus on communication using the foreign language.

The most relevant point of Sauvignon's definition is the reciting of dialogues. As will be shown in the analysis section, it is highly probable that one group of test takers recited dialogues that they themselves created. This definition is important to remember when considering the analysis section of this paper. A definition of washback is next.

Washback refers to various influence on testing and learning (Cheng and Curtis, 2004). The concept originates from the idea that examinations should drive the teaching and learning process. This concept is known as measurement-driven instruction (Popham, 1987). The distinction between test washback and test impact is scope. Wall (1997) defines impact as "any of the effects that a test may have on individuals, policies or practices, within the classroom, the school, the educational system, or society as a whole" (p. 291). There are two types of washback: positive and negative.

Intuitively, many people are fairly familiar with negative washback. Roughly 60 years ago, Vernon (1956) claimed that teachers tended to ignore subjects and activities that did not contribute directly to passing the exam, and that examinations "distort the curriculum" (p.166) ; Wiseman (1961) noted that paid coaching classes, which were intended for preparing students for exams, were practicing test-taking techniques at the expense of learning activities (p.159). Wiseman believed that testing devices had become teaching devices; that teaching and learning were effectively being directed to past exam papers, making the educational experience narrow and uninteresting. The modern phrase that could be used to describe this would be drill and kill. This is where students repeatedly do practice questions during class time to achieve automaticity when presented with such a type of question on an exam.

Smith (1991) in a study on testing in US elementary schools said that "testing programs substantially reduce the time available for instruction, narrow curriculum offerings and modes of instruction, and potentially reduce the capacities of teachers to teach content and to use methods and materials that are incompatible with standardized testing formats"(p.8). These are all examples of negative washback. Let us now move onto positive washback.

Pearson (1988) argues for a mutually beneficial arrangement between learning and testing. Good tests will be more or less directly usable as teaching or learning activities. Similarly, good teaching-learning tasks will be more or less directly "usable for testing purposes, even though practical or financial constraints limit the possibilities"(p.107). In other words, the types of learning tasks that are done in the classroom should also appear in the assessment process. An example of what Pearson refers to is a role play. The role play is used as a teaching device during class time then the students are expected to perform the same role play during the assessment process. Therefore, curriculum alignment (Shepard, 1990) is highly encouraged. The content and the format of the test or examination should mimic the curriculum content. The test is the driver of the curriculum creation process. Testing can have a chilling effect on the kinds of tasks and activities during class time but that is the fault of the test. An ill-conceived test can and will produce negative washback effects. Therefore, the key is to develop better tests. Better tests are those that align with the curriculum and allow the freedom of learning tasks to occur during class time.

The Participants

The participants were 14 first-year Japanese university students who all had six years of English education and could be classified as false beginners. The term has the connotation that the participants have upper-beginner understanding of written English but their communicative ability is low in either spoken or written form. This low communicative ability is the result of six years of English language education that rewards grammar-translation and the various types of summative assessments which are mainly discrete point multiple choice tests. Also, there is a high stakes exam in Japan, called the National Center Test, where the students are not required to communicate but just read,

listen and answer written questions. Very little, if any, communicative competence is required for the examination.

The students had 50 classes of instruction where the emphasis was on oral communication in English. Question forms were taught to the students in previous lessons and by using the questions and answering them, they were expected to engage in conversations almost every class. Having learned the questions and how to use them appropriately and how to use follow up questions gives students the tools to demonstrate some degree of communicative competence. Students were explicitly taught such skills during the term because many second language learning students find carrying on a conversation very hard to do in their first language, let alone a second language.

The strategy that the learners were taught in formula form is Q⇨SA+EI: Q stands for Question, SA stands for Short Answer and EI stands for Extra Information. This means that when someone asks a question, the person being asked should provide a short answer plus extra information. This extra information facilitates follow-up questions and signals that the person is willing to engage in conversation. For example, the question from one test taker to the other could be "Have you ever traveled abroad?" The SA part would be "Yes, I have been to Korea" and the EI part could be "I had a good time and want to go again." The questioner could react to this with a follow up question such as "Oh, really? What did you do there?"

Test Format

A diagram of the test is shown in Appendix 1.0. The ovals represent the two test takers, one oval per test taker. After properly introducing themselves to one another, with a proper handshake, the participants would ask each other questions to come to some point of commonality. Once that commonality has been established, the conversation would continue about that commonality for about four minutes. For example, if the test takers discover they both have a common interest in a particular leisure time activity, like playing a musical instrument, they will talk about playing a musical instrument for four minutes. When about four minutes has elapsed, one of the test takers will politely excuse themselves and end the conversation with a phrase that was learned in class such as "I am sorry, but

I have to go home now." Only saying "bye" would seem unnaturally abrupt in almost any situation and the learners were expressly told not to make such an utterance during class time. The goal of the class was to teach students how to engage in conversations and small talk.

The Intra- Task Variable

Only one variable changed between the two groups. Group 1 test takers chose who their testing partner would be one week before the exam, allowing scripts to be prepared and learned prior to the exam. In other words, participants were able to strategically plan for the exam.

Group 2 test takers did not know until minutes before the exam who would be their partner. Also, their partner was someone that they did not know very well.

Method of Analysis

Conversations contain characteristics. The following characteristic list is provided by O'Sullivan (2008).

Question	Examples
Fillers (F)	Included the use of fillers e.g. uh, OK, um.
Rephrasing (RP)	Paraphrasing the participant's response.
Repetition (R)	The participant repeating their own utterance.
Question Refocus (QR)	No response time given to the participant, the questioner immediately rephrases or redirects the question.
Response	
Minimal Response (MR)	Responding to the questioner with utterances. Such as yeah,mmmmm, uh-huh.

Repetition（R）   The participant repeats the questioner's question.

Clarification Requests（CR）   Where the questioner explicitly requests a clarification of
                                 an utterance.

Expansion（E）   These were question/statements designed to elicit message
                                 expansion which were seen to deviate from the set of
                                 question prompts. e. g. "So what did you do after that?"

Expressions of Interest（EI）   Where the questioners used a phrase such as "Is that
                                 right?" or "That's interesting," or uses intonation to
                                 show marked interest in the response.

Correction（C）   Three types of corrections were expected: lexical usage,
                                 pronunciation, and grammar

Data Collection and Analysis

The total amount of characteristics per pair for each group was tallied. A t-test
analysis was performed and the results are displayed below.

	Group 1	Group 2
Mean	15	25.85714
Variance	13.66667	9.142857
Observations	7	7
Degrees of Freedom	12	
t Stat	-6.0146	
P（T<=t）one tail	3.04E-05	
t Critical one-tail	1.782288	

A t-test is used quite frequently in second language acquisition studies. It is used to examine whether two sample means came from two different populations or not. The results indicate that with more preparation time, the number of speech style characteristics uttered was reduced. Reduced speech style characteristics suggest that the conversations were more scripted.

Weakness of the Study

One of the most obvious weaknesses of the study is the sample size. Given the assumption of a normal population, a larger sample size would definitely bolster the claim that given more preparation time prior to the test, the interlocutors would recite prepared scripts.

A short survey asking the participants how they prepared for the exam would also bolster the claim that the students created scripts outside of class.

Characteristics of the learners need to be investigated when faced with the task of having a five-minute conversation. Learners with either shorter learning experiences, such as junior high school students, or longer learning experiences, such as graduate students, might have different outcomes. The degree of comfort with human interaction of the individual test takers needs to be taken into consideration as well.

The Pedagogical Consequences

The pedagogical consequences are immense for those educators who teach primarily beginners and false beginners. There are educators who condemn audio linguistic style lessons, where students practice and memorize dialogues, especially for lower proficient and false beginners. However, it can be argued that rote memorization is the first step to acquire a second language. Once words and phrases are memorized, the next step for learners is to be placed in situations where they can be used. Placing learners in less structured role plays allows them to demonstrate the degree of communicative competence.

The mere production of language, however, does not mean that learners have achieved communicative competence. Remembering Sauvignon's definition, the ability to use language appropriately and spontaneously is the definition of communicative

competence. Reciting scripted language is not evidence of communicative competence. Therefore, to truly test the degree of communicative competence, less preparation time should be given to the test takers for this kind of assessment.

Like many Japanese learners of English, lack of knowledge of their partner could have resulted in increased anxiety. This possible increase in anxiety in having to talk to someone they are not familiar with is consistent with previous studies of Japanese university students willingness to communicate（Matsuoka, 2009）. The results found in these studies are consistent with research on other types of specific communication anxiety, which states that anxious learners generally speak, write, and participate less in the language classroom than relaxed students（Spolsky 1989）.
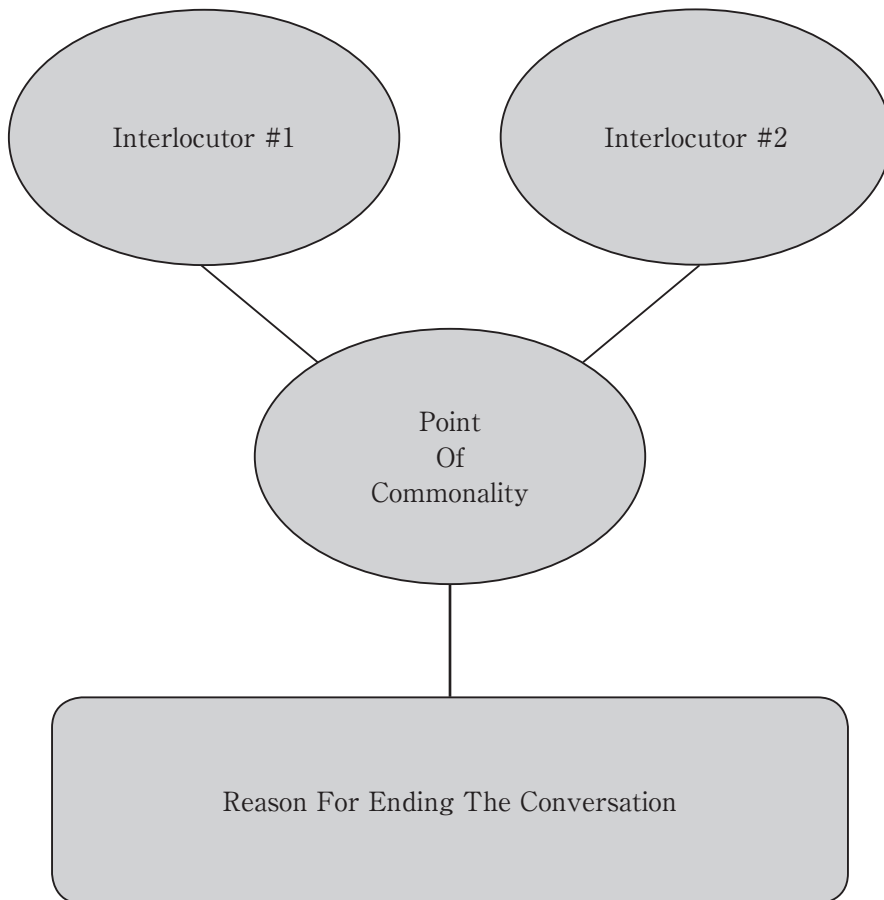
Conclusion

The research question for this paper was: Does increased preparation time lead to more scripted conversations for a five minute conversation test? Two similar groups of learners were analyzed using a list of common natural conversation characteristics. Pairs in Group 1 had more preparation time with their partner than pairs in Group 2. Group 1 significantly uttered less conversations characteristics that appear in natural conversations. This result indicates that test takers prepared by coordinating what they would say to each other prior to the exam.   Therefore, giving learners less preparation time could provide a truer measure of their communicative abilities. Positive washback can facilitate this and still allow for learning tasks to be done in a classroom that emphasize communicative competence without narrowing the teaching techniques of teachers. Therefore, using the test as the driver for curriculum design still can have positive results and lead to learners having improved communicative competence.

Works Cited

Canale, M. and M. Swain. "Theoretical bases of communicative approaches to second language teaching and testing." Applied Linguistics. 1. 1980 1-47. Print.

"Impact and washback in language testing." *Encyclopedia of language and education: Vol. 7 Language testing and assessment.* 1st ed. 1997. Print.

Matsuoka, Rieko. *Japanese College Students' Willingness To Communicate In English.* Saarbrucken, Germany: Lambart Academic Publishing, 2009. Print.

O'Sullivan, B. Modelling Performance in Tests of Spoken Language. Frankfurt: Peter Lang, 2009. Print.

Popham, W. J. "The merits of measurement-instruction." Phi Delta Kappa. 68. 1987: 679-682. Print.

Savignon, S. J. *Communicative Competence: An Experiment in Foreign Language Teaching.* Philadelphia: The Centre for Curriculum Development Inc., 1972. Print.

Shepard, L. A. "Inflated test score gains: Is the problem old norms or teaching the test?" Educational Measurement 9 15-22. Print.

Spolsky, B. *Conditions for Second Language Learning.* Toronto: Oxford University Press, 1989. Print.

Appendix 1.0

A Diagram of A Conversation