An Evaluation of an Elicited Translation Speaking Assessment and Associated Review Activity

Aaron DODS

Abstract

This piece of research examines and evaluates this researcher's use of an *elicited translation activity* to assess learner speaking in a first-year university speaking class. It seeks to evaluate how well the activity fared in terms of his three stated aims for the activity. These aims were (1) to efficiently elicit a sample of focused learner language, (2) to replicate authentic, "real-life" language use outside of the classroom, and (3), for the activity to play an on-going role as a learning activity in its own right. Seeking to evaluate these three aims against the SLA literature, learner feedback and his own observations, the researcher found that the elicited translation activity and its associated review activity performed well. However, although elicited translation did, in a number of key aspects, elicit a sample of focused learner language as well as replicate authentic, "real-life" language use outside of the classroom, it potentially has two significant downsides. Firstly, the lack of a listening component can impact on the authenticity, or cognitive validity, of the speaking assessment. Secondly, L1 transfer, or interference, appeared to result in the elicited sample of learner language digressing from spontaneous language production or other forms of elicited production. Finally, the researcher proposes potential work-arounds for these issues. Firstly, the introduction of delayed translation by the learners, making the assessment more akin to an *elicited imitation* type activity. Secondly, an ad-lib, outcome-based component towards the end of the elicited translation activity that necessarily included listening to the other participant in the conversation and responding accordingly.

1. Introduction

The ability to speak in a foreign language lies at the very core of what it means to be able to use that language well. It is also the most difficult of the four language skills, as its real-time nature demands a lot of the learner in terms of planning, processing, and producing the foreign language. While the speaking of a foreign language is not always the primary goal of foreign language teaching, teachers invariably desire for the L2 to be spoken in the classroom. Additionally, between formal assessments, teachers often evaluate how well a course is progressing on how much the target language is being spoken in the classroom by learners.

Regardless of its importance in most language curriculums, speaking is, nevertheless, also the most difficult language skill to assess reliably. One-on-one interviews with individual learners are not only time-consuming, the interviewer can have considerable power over the interviewee and this may affect the learner's performance. Luoma (2004) mentions several studies that have observed this (see Savignon, 1985; Bachman, 1988; van Lier, 1989; Lazaraton, 1992). An alternative method is paired tasks where the interviewer takes a more passive role and observes two learners, who are required to perform some kind of communicative task (e.g., a role play). As this researcher has found, such tasks can make it difficult to assess speaking for two reasons. Firstly, a learner may not have the opportunity to display his/her speaking skills due to differences in language level or personality with the other participant. Secondly, learners may either consciously or unconsciously constrict their output due to the fact that the activity is assessed.

Dissatisfied with both forms of assessment, this researcher sought to design a new speaking assessment activity. He sought to design an activity that was able to (1) efficiently elicit a quality sample of focused learner language, (2) replicate authentic, "real-life" language use outside of the classroom and, finally, (3) play an on-going role as a learning activity in its own right.

In his activity, learner pairs were required to perform a kind of role-play and translate a Japanese language (L1) conversation into English. This method of eliciting specific language from a learner is known as *elicited translation* (Ellis et al., 2005). In this piece of research, he seeks to classify and evaluate this method of speaking assessment in terms of the relevant second language literature, learner feedback and his own observations. He then seeks to assess the implications of his results and how they impact on the future conduct of this kind of speaking test.

2. Background

From April of this year this researcher has been responsible for developing and delivering a speaking course for first-year university students. The course consists of two classes with a total of 31 students. His faculty has given him the freedom to create and employ his own assessment for the course with the small proviso that 15% of the assessment be based on the department's "stamp card" program (see Talandis, et al., 2011), a program that all teachers in the department are required to implement and seeks to develop learner autonomy in and out of class. Apart from that, assessment is entirely up to the teacher.

The course is based around a prescribed textbook, *Developing Tactics for Listening* (Oxford,

2010) which is used jointly by both the teacher responsible for speaking and the teacher in charge of listening. Six of the textbook's twenty-four units are to be covered in each 15 week semester, and which units, as well the order in which they are to be covered, are decided by the faculty. It is up to both teachers to liaise with each other and ensure that they are covering the same units at roughly the same time.

The textbook is the second level in the three level series *Tactics for Listening* series and, according to the "Series Overview" (p. ix), "is intended for pre-intermediate students who have studied English previously but need further practice in understanding everyday conversational language." The book features twenty-four units featuring several listening tasks relating to some aspect of everyday life such as transportation, neighbours, or campus life. More specifically, each unit contains five listenings- the first three listenings focus on introducing new vocabulary and developing receptive use of language such as listening for gist and listening for detail. The final two listenings seek to move the focus to more productive use of language through a focus on pronunciation and, finally, a dictation task.

As can be seen from the title, the textbook is primarily intended to develop learners' listening abilities. It is therefore incumbent on me, as the speaking teacher, to adapt each unit's core theme, language focus, and vocabulary items into conversational activities that develop the learners' abilities to produce the language presented in the book. For each unit, I create conversation-based activities where learners are required to engage in identical or similar conversations to those in the textbook but at gradually increasing levels of difficulty as they progress through the activity. Eventually, it is hoped that learners will reach the point where they can feel confident in participating in conversations that are similar to those initially presented in the textbook. More significantly, however, it is my hope that learners will have the desire, confidence, and ability to "think on their feet" and use the language they have encountered in the unit to produce their own original contributions to the conversations.

Having largely carte blanche to develop my own course assessment of learners for the course, I had decided on the following; 3x speaking assessment activities (25% each for a total of 75%), learner attitude 10%, and stamp cards (15% as required).

3. Previous task design for the assessment of speaking

In terms of assessing learner speaking, I had envisaged very early on a move away from the open role-play type of paired interview that I had reflexively used in the past where two

examinees were required to carry out a conversation on a topic that had been covered in class. I would then observe and rate learners according to a set of five or six ratings criteria (e.g., variety and accuracy in grammar usage, variety in vocabulary, delivery, level of detail, body language). This kind of task is what Corder (1976; reproduced in 1981) terms as *clinical elicitation* and suggests that it is used "where the investigator has not yet formed any well-formed hypothesis about the nature of the language he is investigating" (p. 69). In this kind of task, learners are oriented primarily toward conveying a message (i.e., fluency) and their responses are termed by Ellis (et al., 2005) as *free constructed responses*.

Over time, however, I had grown more and more aware of the limitations of such an approach. Firstly, differences in the language levels of each participant often led to situations where the examinee with the higher language level was not given sufficient opportunity to demonstrate his or her speaking skills at his or her best. Weir (1993) and Iwashita (1999) observed similar occurrences in their own research. Secondly, even learners who had demonstrated sound speaking ability in similar conversations in class resorted to reducing their language output in a similar assessed conversation. This act, where learners adapt what they say according to their language resources, is termed by Bygate (1987) as a *reduction strategy* and this researcher considers both factors as having had a significant detrimental effect on the quantity and quality of the sample of learner language that he has been able to elicit.
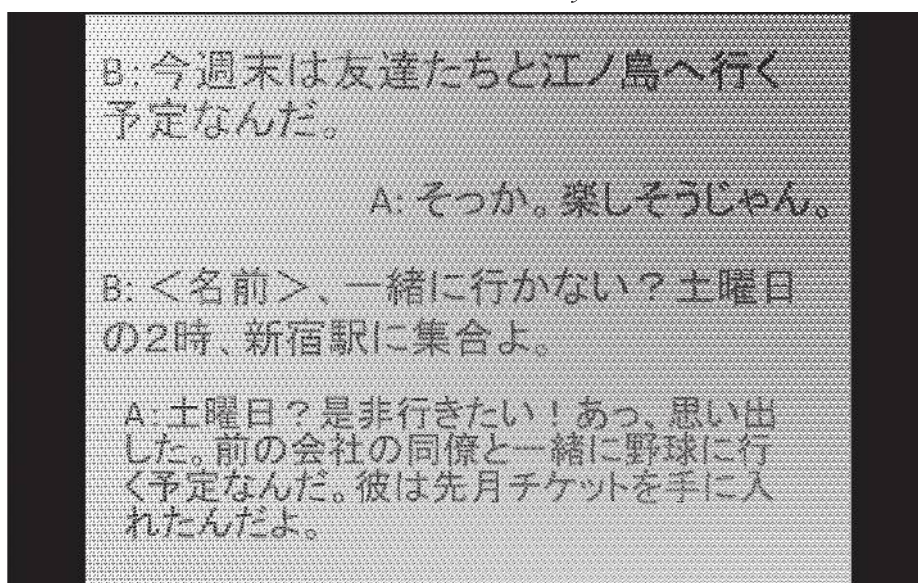
The assessment activity

B：今週末は友達たちと江ノ島へ行く
予定なんだ。

A：そっか。楽しそうじゃん。

B：＜名前＞、一緒に行かない？土曜日
の2時、新宿駅に集合よ。

A：土曜日？是非行きたい！あっ、思い出
した。前の会社の同僚と一緒に野球に行
く予定なんだ。彼は先月チケットを手に入
れたんだよ。

Figure 1: Screenshot of slide from Speaking Assessment Activity

4. A new task design for the assessment of speaking

Firstly, I wanted to design a task that collected a larger sample of learner language that better reflected learner speaking ability as well as elicit specific language items that had been covered in the classroom. Secondly, I wanted the task to, as best as possible, replicate the use of language in real-life communicative situations outside of the classroom. Finally, I wanted the activity to have the ability to act as not only an assessment activity but also as a stand-alone activity that contributed to on-going language learning.

In the elicited translation activity, learners are required to perform a "conversation" by translating Japanese language (L1) sentences into English (L2). Two learners, who have been told well in advance which topics from the textbook are to be assessed, are randomly selected to stand in front of a large projector screen. The screen is used to display the Japanese language conversation (i.e., the input) via the presentation software PowerPoint. Once each learner has been assigned to either role "A" or role "B," "turns" in a Japanese language (L1) conversation incrementally appears before the learners (see Figure 1). Here, a "turn" represents the time and space in a conversation where one person starts and stops speaking. A conversation consists of two or more participants taking turns to speak, or "turn-taking," and may even include "overlapping turns." As either A's or B's turn appear on the screen, learner A or B is required to produce the L1 language as best as they can in the L2 (English, the "output"). If the learner hesitates or pauses for an unnaturally long period of time, the teacher presses the space bar on the keyboard and move on to the other learner's turn. This process is carried out for two different conversations on two different topics that were covered in class in order to ensure that both learners have the opportunity to initiate a conversation as well as receive equal amounts of speaking time.

Rating Criteria

My criteria for rating learner performance was developed intuitively and based on four distinct criteria; delivery (clarity, variety in pacing, emphasis), sentence construction (complexity and accuracy), vocabulary recall (accuracy and pronunciation), and formulaic language (ability to appropriately and accurately use commonly-used chunks of language).

A maximum of 5 points was given for each criteria item with 1 signifying "poor," 5 equivalent to "excellent" and 2,3,4 somewhere in between. As a result, 20 constituted a maximum score for each of the two assessments, with a total maximum assessment score for both assessments of 40 points. For each learner, I used an L2 transcript of the conversation and, in real-time,

gave each turn a rating from 1-5 for the accuracy-related criteria of sentence construction and vocabulary recall. I gave a rating for the third accuracy-related criterion of formulaic language for only the three or four turns that required it. Finally, I gave an overall score for delivery at the end of the assessment. After the activity, I calculated the average rating for each of the three accuracy-related criteria, added the rating for delivery and presented each learner's final score for each conversation at the bottom of the transcript. This transcript was then handed out to each learner in the following lesson as feedback after the review activity to be filed with theis other class handouts.

Review Activity

In the following lesson, once learners had received and looked through their feedback, I quickly went through the PowerPoint speaking assessment in order to refresh their memories of the conversation situations and language that had been elicited. Then, I introduced learners to a review activity where learners were given time to discuss with other learners a potential L2 translation for each L1 turn per slide. Finally, at the end of each slide, learners were shown a model L2 translation for each L1 turn (see Figure 2) and time was allotted for learners to discuss these translations with their partner or myself. This review activity is also intended to be used at other opportune moments during the course such as in the final or first classes of the second semester.
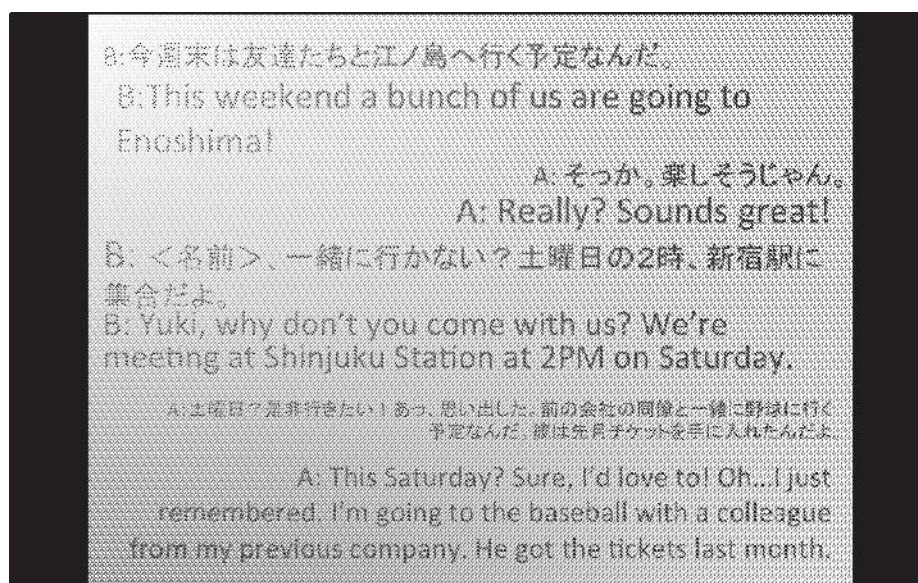

Figure 2: Screenshot of Review Activity (same slide as in Figure 1)

5. Classification and evaluation of the speaking assessment activity

This elicited translation seeks to collect not clinically elicited samples of learner language as in an open role-play, but, rather what Ellis (et al., 2005) calls *experimentally* elicited samples. In contrast to a clinical elicitation, Ellis (et al.) states that an experimental elicitation seeks to collect "specific, pre-determined linguistic feature[s] in learners' production" (p. 23). It focuses more on form (i.e., accuracy) than message conveyance or fluency and seeks to elicit *constrained constructed responses* (Norris and Ortega, 2001) rather than free constructed responses.

This assessment activity would appear to be a hybrid type of elicited translation/role-play activity. The elicited translation aspect of the activity seeks to guide learners in the language forms to be produced and requires them to focus on producing specific, pre-determined language. The role-play aspect of the activity intends to simulate the use of language in real-life communicative situations and is intended to imbue the activity with what is termed *cognitive validity* (Field, 2011), that is, it enables the activity to better resemble the cognitive process of planning, processing, and producing language in an authentic, real-time conversation.

How well does the elicited translation activity elicit a sample of focused learner language?

The translation nature of the speaking assessment activity allowed me to select and attempt to elicit specific structural, lexical, and formulaic aspects of the L2. However, Ellis (et al., 2005) notes that there is considerable disagreement within the SLA community as to how valid clinically elicited or experimentally elicited data actually is. While some researchers (Naiman, 1974) argue that elicited data closely matches naturally occurring language, others have observed a number of differences between elicited speech and spontaneous speech. Schumann (1978) observed that a learner he was investigating produced markedly different types of negative utterances in spontaneous speech (*no* + verb) as compared to elicited speech (*don't* + verb). Burmeister and Ufert (1980) carried out a careful comparison of data collected from spontaneous speech and translation tasks and observed that a number of non-target like structures that never or rarely occurred in spontaneous speech were more frequent in the elicited data. Ellis (et al., 2005), in referring to *elicited translation*, states that if the sentences to be translated into the L2 are long enough, the learners will be required to "re-encode the meaning using their own linguistic resources" (p. 38). While this is a perfectly valid method to assess a learner's ability in L2, Ellis warns that the danger of such elicited translation activities is that the transfer of language items and structures that are not the same in L1 to L2, otherwise known as *negative L1 transfer or*

interference, may occur when in more natural language use it would not (see Burmeister et al., 1980; Lococo, 1976).

Does the activity replicate "real-life" language use outside of the classroom?

Bygate, in his model of speech (1987), considers the uniqueness of speaking as compared to the other three basic skills as being due to *processing* and *reciprocity*. In terms of processing, speaking requires simultaneous action. As Bygate states, "the words are being spoken as they are being decided and as they are being understood" (p. 11). As for the reciprocity condition, it refers to the speakers having to adapt to their listeners and adjust what they say according to their listener's or listeners' reactions.

In terms of the first condition, processing, elicited translation certainly features some degree of cognitive validity. The learners do not know what they are going to be required to say until their turn appears; they need to start their turn within a few seconds of reading the L1 on the screen. This is similar to a natural conversation (in either L1 or L2) where a participant is not able to plan what he/she is going to say until another participant has begun or finished their turn. The difference is that in a natural conversation a participant decides what they are going to say after having listened to another participant. However, in elicited translation, since what is going to be said has been pre-determined by the teacher for purposes of assessment, listening to the speaker is not required. Elicited translation, then, appears to satisfy the processing condition for cognitive validity, although in a clearly artificial manner; the need to listen to another participant in the conversation is eliminated.

Accordingly, elicited translation appears to function well as a stand-alone activity assessing just one skill (i.e., speaking) rather than an integrated-type activity that assesses more than one skill (i.e., both speaking and listening). Integrated assessments are often able to reflect authentic, natural language use and therefore may be evaluated highly in terms of cognitive validity. While elicited translation, as it assesses only the one skill of speaking, is not be an integrated-type activity, it may be that since only listening skills are assessed in the listening class (taught by another teacher), perhaps only speaking skills should be assessed in my own speaking class.

In terms of the second condition, *reciprocity*, cognitive validity may be considered to exist, but once again, only in a somewhat artificial, circuitous fashion. There is no need to listen to another participant as participants are not required to adjust what they say according to the other

participant's or participants' reactions; they simply say what appears on the screen when it is their turn. However, some form of cognitive validity is surely provided by the fact that (1) participants cannot know what they are going to say until it is actually their turn and (2) what a participant might be required to say may differ from what they imagine they will be elicited to say as a result of deliberate adjustments having been made to the assessed conversations (as compared with the conversations carried out in class).

Are the rating criteria appropriate, workable?

Production accuracy is assessed in the three criteria of sentence construction, vocabulary recall, and formulaic language. The only criterion that does not assess accuracy in production is delivery. Delivery, nevertheless, is an integral component of any speaking test as such factors as volume, pacing, and pronunciation impact just as much, if not more, on the ability to speak well as does lexical and grammatical accuracy. This emphasis on accuracy would appear to be consistent with the focus on form or accuracy that is afforded by the elicitation of specific, or as Ellis (et al., 2005) calls them "pre-determined" (p. 36) linguistic features afforded by the clinical elicitation of learner production.

Formulaic language was specifically included as a criterion because of this researcher's belief in its importance in enabling fluid, grammatically, and meaningfully accurate speech. It is an umbrella term referring to prefabricated chunks of language that encompasses much more than just idioms and collocations (Wray, 2000). Their use by native speakers and L2 learners has been demonstrated to aid fluency as they require less processing power than creating sentences from scratch (see Kuiper, (2004); Nesselhauf, (2005); Schmitt & Carter, (2004); Wray, 2000, 2002). Examples of formulaic language that were elicited in the speaking assessment activity included idioms such as ("take a rain check"), collocations ("go out"), sentence frames ("Why don't you...?"), fluency devices ("how can I say," "you know") and politeness markers ("Can I ask...").

In terms of the workability of actually using the criteria in real-time, assigning several ratings to each learner turn while at the same time needing to push the space bar to move to the other learner's turn proved taxing. Possible solutions might include setting the next turn to appear a specified amount of seconds after the previous turn or using a Bluetooth presentation clicker in my left hand to move to the next turn while assigning the ratings for each turn with my right hand. Additionally, I had virtually no time to take notes of samples of learner production. Doing so has merits not only in terms of giving feedback to learners but also in terms of

collecting samples of L1 transfer for further research. A potential solution might be to make an audio recording of each conversation and then go through conversations that I have flagged at a later date. However, this would require written consents to be completed by each student and also be time-consuming to go through each conversation.

How well can the elicited translation activity and subsequent review activity play a role in on-going learning?

The elicited translation speaking assessment activity may itself have the potential to play a role in learning. The review activity, which is based on the speaking assessment activity, should also have benefits for learners.

In the elicited translation, learners are re-encountering L1 words that they originally met and understood the meaning of in class via vocabulary list completion activities, word card memory recall activities and finally, productive language activities such as role plays. This re-encountering of words may help in the remembering of L2 words through the concept of *retrieval* (Baddley, 1990).

Retrieval is where learners initially notice a word, comprehend its meaning, and subsequently need to retrieve that word. It is important to note that the process of retrieval cannot occur if the target language word and its L1 translation are presented simultaneously. Retrieval can be divided into two types, receptive and productive. Receptive retrieval involves noticing a language item and retrieving its meaning when the word is encountered during listening or reading. Productive retrieval involves the desire to communicate the meaning of a word and having to retrieve its spoken or written form during speaking or writing.

In the elicited translation, learners are seeking to retrieve a word and then produce its spoken form, which represents productive retrieval, the most cognitively taxing form of retrieval. The fact that learners are seeking to productively retrieve words, grammar and formulaic language within the context of an activity that possesses a high degree of cognitive validity surely contributes to the on-going learning and remembering of that word.

With respect to the review activity, learners are once again presented with the L1 turns and language, given time to discuss possible L2 translations with other learners, and finally, are presented with potential L2 translations for each L1 turn per slide. This is repeated for all slides.

Such an approach not only has merits in terms of the SLA concept of retrieval, but also in terms of *negotiation* and *noticing*.

Negotiation has been shown to increase the likelihood of words being learned (Ellis, 1994; Newton, 2013). Discussion amongst learners of potential target language equivalents before being presented with a potential translation can be considered a form of negotiation, as can questioning by learners of the teacher's presented translation. A significant feature of classroom negotiation that studies have revealed is that learners who observe negotiation can learn vocabulary just as well as learners who do the actual negotiation (see Ellis and Heimbech, 1997; Ellis, 1994; Newton, 2013; Stahl and Clark, 1987; Stahl and Vancil, 1986). This would seem to indicate that it is not the negotiation itself which is important but, rather, a collaborative environment where interpretation of meaning or translations of facets of language is being exchanged between learner and learner, teacher and learner, and learner and teacher.

The SLA concept of *noticing* (see Ellis, 1990) is where learners notice an aspect of language (whether it be a word, grammar or formulaic language) within a written or oral language input and become aware of it as a potentially useful language item. How well a learner notices a particular language item can be affected by such factors as its salience within the textual input, previous contact that learners have had with it, and learners' realization that the item fills a gap in their knowledge of the language (Ellis, 1990; Schmidt and Frota, 1986).

The final stage of the revision activity is where learners, who have just negotiated what they feel is the best L2 translation are presented with what the teacher sees as a good translation of the item, provides significant opportunities for noticing. Where the learner's translation is different to that presented, it is possible that learners will notice a gap and, consequently, learn the language item. Of course, how well they learn it will depend on the three factors outlined above. However, given the fact that the item comes from an assessment activity consisting of language that learners have had previous contact with, as well as the fact that not knowing the language item reflects a gap in their knowledge that impacts on their score in an assessment, there seems to be considerable potential for learning through noticing. Similarly, where the learner's translation of the item was correct, this likely serves to embed the item within the learner's memory.

6. Learner evaluation of the elicited translation speaking assessment and review activity with researcher responses

A questionnaire containing thirteen multi-choice questions and three questions where learners could fill in their own comments was distributed to learners who filled it in during class time. For the multi-choice questions, learners had the choice of checking "yes, absolutely," "yes, somewhat," "no, not really," or "no, not at all." Questions were presented in both English and Japanese, and learners had the option to write any answers in either Japanese or English. A total of 30 anonymous questionnaire forms were received from the learners and the results were analyzed in Microsoft Excel.

Learners overwhelmingly considered the assessment activity as a good way to assess speaking ability with 100% answering either "absolutely" (15/30) or "yes, pretty much" (15/30) to the question "Do you think this kind of speaking test is a good way to understand how well you can speak English?" More specifically, most learners considered the assessment activity as testing language covered in class with 77% (23/30) agreeing that it tested studied vocabulary, 80% (24/30) that it tested studied grammar and 90% (27/30) that it tested studied formulaic expressions. Somewhat less emphatically, only 33% (10/30) indicated they enjoyed the test and 60% thought they could do well in the test if they studied harder. These two results may reflect learner anxiety about speaking a foreign language that they have had few opportunities to use in their daily lives.

However, issues relating to the fairness of the assessment arose with 33% (10/30) answering "not really" to the question "Do you think the test was fair?" In explaining why, four learners indicated that they felt the degree of difficulty between roles A and B differed. Another learner indicated that she felt that one of the roles simply required the use of short pieces of formulaic language while the other role required lengthier full sentences. Additionally, three learners indicated that learners who took the assessment later than others were given an unfair advantage because those learners who had already taken the test talked about it with their friends upon returning to the classroom. With respect to the first issue, I need to pay more attention to ensuring that both roles are roughly equal in terms of what kind, and how much, language is being elicited. In terms of the second issue, although I took the precaution of telling each pair of learners who had just taken the test to not mention it to learners who had still to take it, there was invariably discussion amongst learners of the test's content and how they think they performed in it. A solution would be to have learners who have finished the test move to another

area such as an empty classroom or the English lounge area.

In terms of the stated first aim of this assessment activity, effectively eliciting a quality sample of focused learner language, an issue that was raised by a few learners was that they found some of the L1 (Japanese) turns too long. This correlates with this researcher's observation that, in cases where there was a significant amount of information that needed to be conveyed, there was a tendency amongst some learners to translate word-by-word, sentence-by-sentence, without any consideration of what the overall "idea" of the turn was intended to be. As a result, there appeared to be cases of negative L1 transfer in learners' production and the opening of an apparent gap between spontaneous language use and elicited language. SLA research (see Burmeister, et.al, 1980; Lococo, 1976) has found that L1 transfer can be a potential downside of translation-based tasks.

There might be a few potential solutions to this issue, one in the classroom and two in the conduct of the assessment. Firstly, in the classroom, I might need to better emphasize to learners that speaking differs from writing in that, rather than consisting of complete sentences, it consists of simpler "idea units" (Luoma, 2004). These idea units are made up of short phrases and clauses that are connected with *and*, or, *but*, or *that*. Sometimes, instead of being connected by words, phrases, and clauses, they simply have a short pause between them. Secondly, in the actual conduct of the test there may be merit in showing each L1 turn to the learner for a few seconds and then having him/her begin translating into L2 only once the turn has been removed from the screen. Such an approach makes the activity more analogous to an *elicited imitation* type activity (as opposed to elicited *translation*) in that the learners will not be able to remember the exact sentence(s) but, rather, have to imitate it by first processing it for meaning and then "re-encoding it using their linguistic resources" (Ellis, 2005; p. 38). Doing so would surely reduce word-by-word or sentence-by-sentence translation and any resulting L1 transfer, not to mention encourage the learners to focus on conveying the actual idea of the turn through use of idea units. Another potential solution to the issue of L1 transfer is adding an ad-lib segment at the end of each task where each learner is given a specific goal in terms of how they want the conversation to end. Doing so may not only reduce L1 transfer, it would add to the cognitive validity and integrated nature of the activity by creating the need to listen to the other participant.

In terms of the second aim of this assessment activity, replicating authentic, "real-life" language use outside of the classroom, 93% (28/30) of learners indicated that they found the test to be difficult. This corresponds with final learner scores for the assessment where the averages

for each class that took the test were 55% and 56% respectively. Eight learners out of the fifty-one (16%) who took the assessment obtained scores of 75% or more and, at the other end of the scale, ten learners (19%) received scores of less than 40%. Where reasons were given, they could generally be differentiated as follows;

· converting the sentences into English within a limited time

· recalling the grammar and vocabulary that had been studied

· knowing words but not being able to "properly" put them into sentences

· not being able to translate or express a word in the test that did not appear in the textbook's vocabulary list.

· not knowing how to fill hesitations in production

These issues are ones that learners would encounter in an authentic, real-time L2 conversation. Confirming this interpretation of responses, 90% (27/30) of learners said that the speaking test felt like a real conversation. It would seem reasonable to conclude, then, that, even without a listening dimension, the speaking assessment reasonably achieved this researcher's second aim of replicating a real-time, authentic L1 conversation; it possessed a not unsubstantial degree of cognitive validity.

Finally, in terms of the third aim of this assessment activity, playing an on-going role as a learning activity in its own right, 83% (25/30) of learners considered the test itself as being good practice for improving their English and 100% (30/30) considered the review activity as being a useful activity for improving their English. While most learner comments were related to speaking tests in general (e.g., "don't get many opportunities to speak English," "best way to speak English is to speak it") a couple of learners alluded to the fact that, they felt they could "acquire practical and applicable skills required for speaking" through the test. Another learner noted a change in terms of focusing more on communicating meaning and not worrying about linguistic mistakes. A couple more learners mentioned the informal, conversational aspect of the two tasks and said that it might help them when they go overseas or do business using the L2. Speaking specifically about the review task, one learner mentioned that he enjoyed understanding the differing nuances between what was imagined by a learner as being a good L2 translation and what the teacher recommended as being more natural English. In terms of improvements, two learners mentioned that they wanted to receive the review activity as a handout featuring both the L1 and L2 recommended translations at the completion of the activity.

7. Conclusion

In terms of achieving this researcher's aims of (1) eliciting a quality sample of focused learner language, (2) replicating authentic, "real-life" language use and, (3) playing an on-going role as a learning activity in its own right, evaluation of the speaking assessment activity and associated review activity against the SLA literature and learner feedback demonstrated a more than satisfactory performance, especially in terms of the third aim.

Nevertheless, SLA literature, learner feedback, and researcher observations also revealed two issues related to the fact that learner language production is elicited through translation. Firstly, no listening is required on behalf of learners and, as a result, the ability of the activity to replicate genuine spoken language, or cognitive validity, is diminished. The researcher counter-ed this by raising the point that, as this is a speaking class and learners attend another class focusing on listening, it might be better to not have learner receptive ability potentially interfer-ing with learner productive ability in the speaking class. Secondly, it is highly likely that translation-elicited production results in negative L1 transfer, or interference, and as a result, leads to language production that is not representative of the learner's real speaking ability. In order to counter this, the researcher suggested having learners begin translation after the L1 language has been removed from the screen. He also raised the possibility of adding an ad-lib outcome-focused component at the end of each task that would not only potentially reduce L1 transfer but also improve the cognitive validity of the test by adding a small listening component to the assessment.

References

Bachman, L.F. (1988). Problems in examining the validity of the ACTFL oral proficiency interview. *Studies in Second Language Acquisition 10*, 2, p.149-164.

Baddley, A. (1990). *Human Memory*. London: Lawrence Erlbaum Associates.

Burmeister, H. and Ufert, D. (1980). 'Strategy switching' in S. Felix (ed.): *Second Language Development*. Tubingen: Gunter Narr.

Bygate, M. (1987). *Speaking*. Oxford: OUP.

Corder, S.P. (1976) 'The study of language' in proceedings of the Fourth International Conference of Applied Linguistics. Munich, Hochschulverlag. Also in Corder (1981). *Error analysis and Interlanguage*. Oxford: Oxford University Press.

Ellis, R., (1990). *Instructed Language Learning*. Oxford: Blackwell.

Ellis, R., (1994). Factors in the incidental acquisition of second language vocabulary from oral input: a review essay. *Applied Language Learning*, *5*, 1, p.1-32.

Ellis, R., and Barkhuizen, G., (2005). *Analysing Learner Language*. Oxford: Oxford University Press.

Ellis, R., and Heimbech, R. (1997) Bugs and Birds; Children's acquisition of second language vocabulary

through interaction. *System, 25*, p.247-259.

Field, J. (2011). Cognitive Validity in Taylor, L. (ed.): *Studies in Language Testing*. 30. Cambridge: Cambridge University Press.

Iwashita. N. (1999). The validity of the paired interview format in oral performance assessment. *Melbourne Papers in Language Testing 8*, 1, p.51-66.

Kuiper, K. (2004). 'Formulaic performance in conventionalised varieties of speech' in N. Schmitt (ed.). *The acquisition, processing, and use of formulaic sequences* (p.37-54). Amsterdam: John Benjamins Publishing Company.

Lazaraton, A. (1992). The structural organization of a language interview: a conversation analytic prospective, *System 20*, p.373-386.

Lier, L. van (1989). Reeling, writhing, drawling, stretching and fainting in coils: oral proficiency interviews as conversation. *TESOL Quarterly*, p.489-503.

Lococo, V. (1976). 'A comparison of three methods for the collection of L2 data: Free composition, translation and picture description.' Working Papers on Bilingualism *8*, p.59-86.

Luoma, S. (2004). 'Developing speaking tasks' in *Assessing Speaking* (p.139-169) Cambridge: CUP.

Naiman, N. (1974). 'The use of elicited imitation in second language acquisition research.' *Working Papers on Bilingualism 2*: p.1-37.

Nesselhauf, N. (2005). *Collocations in a learner corpus*. Amsterdam: John Benjamins.

Newton, J. (2013). Incidental vocabulary learning in classroom communication tasks. *Language Teaching Research, 17*, 3, p.164-187.

Norris, J. and Ortega, L. (2001). 'Does type of instruction make a difference? Substantive findings from a meta-analytic review' in Ellis (ed.): *Form - Focused Instruction and Second Language Learning*. Malden, MA: Blackwell.

Richards, J.C. (2010). *Developing Tactics for Listening*. Oxford: Oxford University Press.

Savignon, S. (1985). Evaluation of communicative competence: the ACTFL proficiency guidelines. *The Modern Language Journal 69*, p.129-134.

Schmidt, R.W. and Frota, S. (1986). 'Developing basic conversational ability in a second language: A case study of an adult learner of Portuguese.' in Day, R. (ed.), *Talking to Learn: Conversation in Second Language Acquisition* p.237-326. Rowley, MA: Newbury House.

Schmitt, N., & Carter, R. (2004). Formulaic sequences in action: An introduction. In N. Schmitt (ed.), *The acquisition, processing, and use of formulaic sequences* (p.1-22). Amsterdam: John Benjamins Publishing Company.

Schumann, J. (1978). *The Pidginization Process: a Model for Second Language Acquisition*. Rowley, MA: Newbury House.

Stahl, S.A. and Vancil, S.J. (1986). Discussion is what makes semantic maps work in vocabulary instruction. *The Reading Teacher, 40*, 1, p.62-67.

Talandis, G., Taylor. C, Beck, D., Hardy, D., Murray, C., Omura, K., and Stout, M.(2011). The stamp of approval: motivating students toward independent learning. *Bulletin of Toyo Gakuen University*, p.165-182.

Underhill, N. (1987). *Testing Spoken Language, a handbook of oral testing techniques*. Cambridge: Cambridge University Press.

Weir, C. (1993). *Understanding and Developing Language Tests*. New York: Prentice Hall.

Wray, A. (2000). Formulaic sequences in second language teaching: Principles and practice. *Applied Linguistics, 21* (4), p.463-489.

Wray, A. (2002). *Formulaic language and the lexicon*. Cambridge: Cambridge University Press.

Appendix A. 〔Student Questionnaire〕

Please circle the most appropriate answer or write a short answer in the space provided. You can answer in English or Japanese. Thank you!

1. Are you male or female?　　あなたは女性ですか、男性ですか？

    male（男性）　　　　　　　　　　　female（女性）

2. What is your mother tongue?　あなたの母国語は何ですか？

   Japanese（日本語）　　*Chinese*（中国語）　　*Korean*（韓国語）　　*Other* その他（　　　　　）

3. Do you think the speaking tests included vocabulary studied during the course?
　スピーキングテストではコースの学習した単語が多く出てきましたか？

   yes, lots（はい、とてもそう思う）　　　*yes, quite a bit*（そう思う）
   not really（そうは思わない）　　　　　*no*（思わない）

4. Do you think the speaking tests included grammar studied during the course?
　スピーキングテストではコースで学習した文法が多く出てきましたか？

   yes, lots（はい、とてもそう思う）　　　*yes, quite a bit*（そう思う）
   not really（そうは思わない）　　　　　*no*（思わない）

5. Do you think the speaking tests included formulaic language studied during the course?
　(e.g., That's too bad., Can I ask...,　I'm tired of...)
　スピーキングテストではコースで学習した定形表現が多く出てきましたか？

   yes, lots（はい、とてもそう思う）　　　*yes, quite a bit*（そう思う）
   not really（そうは思わない）　　　　　*no*（思わない）

6. Do you think that if you study hard, you can do well in the speaking test?

学習したことがそのままスピーキングテストに正しく反映されていると思いますか？

yes, absolutely（はい、とてもそう思う）　　*pretty much, yes*（はい、そう思う）

not really（それ程でも無い）　　　　　　*no*（いいえ、全く思わない）

7. Do you think the speaking tests were difficult? スピーキングテストは難しいと感じましたか？

yes, absolutely（はい、とてもそう思う）　　*pretty much, yes*（はい、そう思う）

not really（それ程でも無い）　　　　　　*no*（いいえ、全く思わない）

8. What was the most difficult thing about the speaking test?

スピーキングテストで最も難しいと感じたことは何ですか？

_____

_____

_____

_____

9. Do you think the speaking tests were fair? If not, why not?

スピーキングテストの環境は公平であったと思いますか？『いいえ』の場合なぜですか？

yes, absolutely（はい、とてもそう思う）　　*pretty much, yes*（はい、そう思う）

not really（そう思わない）　　　　　　　*no*（いいえ、全く思わない）

_____

10. Do you think the speaking tests were similar to having a conversation in English?

スピーキングテストは英会話そのものと似ていると思いますか？

yes, absolutely（はい、とてもそう思う）　　pretty much, yes（はい、そう思う）

not really（そう思わない）　　　　　　　no（いいえ、全く思わない）

11. Did you enjoy the speaking test? Why?　　　スピーキングテストを楽しみましたか？

　　yes, absolutely（はい、とてもそう思う）　　*pretty much, yes*（はい、そう思う）

　　not really（そう思わない）　　　　　　　　*no*（いいえ、全く思わない）

　　_____

　　_____

12. Do you think this kind of speaking test is a good way to understand how well you can speak
　　English? Why?

　　このような英語のスピーキングテストは、自身の英語力を計るのに有効だと思いますか？

　　yes, absolutely（はい、とてもそう思う）　　*pretty much, yes*（はい、そう思う）

　　not really（そう思わない）　　　　　　　　*no*（いいえ、全く思わない）

　　_____

　　_____

13. Were you surprised by performance in the speaking test? Why?

　　このスピーキングテストにより、自分の英語力の想定と結果に相違はありましたか？

　　yes, absolutely（はい、とてもそう思う）　　*pretty much, yes*（はい、そう思う）

　　not really（そう思わない）　　　　　　　　*no*（いいえ、全く思わない）

　　_____

　　_____

14. Do you think the speaking test is good practice for improving your speaking? Why?

　　スピーキングテストはあなたの英語の上達に有効の手段だと思いますか？なぜですか？

　　yes, absolutely（はい、とてもそう思う）　　*pretty much, yes*（はい、そう思う）

　　not really（そう思わない）　　　　　　　　*no*（いいえ、全く思わない）

　　_____

　　_____

15. Do you think the review activity (where the teacher shows the English that he hoped you would say) was a useful activity for improving your speaking?

単語や言い回しに対し、講師が修正や補足をすることは、スピーキング力向上に有効だと思いますか？

    yes, absolutely（はい、とてもそう思う）   *pretty much, yes*（はい、そう思う）

    not really（そう思わない）   *no*（いいえ、全く思わない）

16. Does the speaking test motivate you to study English harder?

スピーキングテストは英語の学習のモチベーションに繋がりますか？

    yes, lots（はい、とてもそう思う）   *yes, a bit   not really   no*

17. Do you think I can improve the speaking test? If so, how?

講師がスピーキングテストを改善できるとしたらどのような点だと思いますか？

_____

_____

_____

_____

18. Any other comments about the speaking test?

その他、スピーキングテストに対し、何かコメントがあればこちらにご記入下さい。

_____

_____

_____

_____