

Bridging the Gap Between Instruction and Assessment: Examining the Role of Dynamic Assessment in the Oral Proficiency Skills of English-as-an-Additional-Language Learners

Jeness Weisgerber *

The University of Victoria
jenessodel@gmail.com

Abstract

This exploratory study investigated the role of dynamic assessment (DA) in improving the oral proficiency skills of English-as-an-additional-language learners. It focused specifically on speaking test scores and the use of language learner strategies, with the goal of providing empirical evidence as well as pedagogical recommendations. Seven participants were administered a section of the IELTS™ Speaking test in both dynamic and standardized formats. Each test was followed by a think-aloud protocol in order to ascertain participants' thoughts and strategic behaviours during the testing process. In terms of test scores, results showed no holistic differences, but did show differences in fluency, grammatical range, and lexical resource scores. Scores for grammatical range and lexical resource were higher in DA, while scores for fluency were higher in standardized assessment. An analysis of the participants' strategic behaviours also showed a greater use of cognitive and metacognitive strategy use in DA. These results point to DA's potential to facilitate the development of grammatical and lexical abilities as well as to foster the use of language learner strategies within the sample.

Keywords: Speaking test; dynamic assessment; English-as-an-additional language; language learner strategies

I. INTRODUCTION

Second-language assessment tools are not only used to assess language skill, but are also perhaps one of the most fundamental learning and teaching tools (Huang, 2014). Increasingly, there is a dichotomy in the field of education between two types of assessment: standardized and dynamic. During standardized assessment (SA), sometimes referred to as traditional assessment, learners receive a set of items or problems and attempt to solve these with minimal or no feedback (Sternberg & Grigorenko, 2002). Conversely, during dynamic assessment (DA), learners receive intervention during assessment, often in the form of feedback (Sternberg & Grigorenko, 2002). DA, in contrast with SA, has a premise based on the role of mediation that “enables learners to perform beyond their current level of functioning, thereby providing insights into emerging capabilities” (Poehner & Lantolf, 2013, p. 323). DA has become of particular interest in the fields of Applied Linguistics and Second Language Acquisition (SLA), due to the dynamic nature of language learning. Many researchers (e.g. Poehner, 2008; Sternberg & Grigorenko, 2002) advocate for DA practices in the classroom. Despite this, according to Poehner (2008), there is a gap in the DA literature within the field of SLA in regards to second language (L2) performance from

*This research was funded by the Jamie Cassels Undergraduate Research Award (JCURA) through the Vice President Academic and the Learning and Teaching Centre at The University of Victoria. I would like to especially thank my supervisor, Dr. Li-Shih Huang, for her continued support, guidance, and mentorship throughout this project. I would also like to thank Catherine Chao for her invaluable assistance with the video-stimulated recall, test rating, and transcribing.

a DA perspective. In an attempt to fill this gap, the current research focuses on examining the role of DA in facilitating the oral proficiency skills of English-as-an-additional-language learners. This exploratory study aims to provide pedagogical insights as well as empirical evidence to help instructors make informed decisions about the best method of classroom-based assessment for improving English language speaking proficiency. It is important to note that the aim of this study is to determine the best method of assessment solely in regards to the goal of improving language proficiency, rather than other potential goals of language assessment (i.e. predicting learner success in a certain English class).

In the sections that follow, I first provide an overview of the key concepts and research in the areas of DA, including the role of language learner strategies. Then, I outline the methods used throughout the current study. Lastly, I present the research results and discuss their implications, particularly in regards to pedagogical recommendations.

II. LITERATURE REVIEW

I. Dynamic Assessment

DA has its basis in L.S. Vygotsky's (1978) notion of Sociocultural Theory (SCT) and the Zone of Proximal Development (ZPD). SCT can be understood as the idea that humans develop uniquely through their interactions with others and their environment, including in the form of mediation using various tools (physical, cultural, symbolic, etc.), and that this leads to higher forms of cognition that would not be possible without this interaction (Poehner, 2008). From this, Vygotsky (1978) defined the ZPD as "the distance between the actual developmental level as determined by independent problem solving and the level of potential development as determined through problem solving under adult guidance or in collaboration with more capable peers" (p. 86). The ideas of mediated learning and skills that are still in the process of developing are central to SCT and the ZPD, and are also core tenants in the concept of DA. Moreover, DA is founded in Vygotsky's influential idea that in the ZPD "instruction leads development" (Lantolf & Thorne, 2006, p. 327). These notions are crucial to understanding and operationalizing DA.

DA is commonly divided into two general sub-types: interventionist and interactionist (Lantolf & Poehner, 2014). Interventionist DA is similar to some methods of SA because it retains certain standardized procedures and has a greater emphasis on results that can be measured and used to compare between tests, both in regards to tests taken by the same learner at different times and between learners (Poehner, 2008). On the other hand, during interactionist DA, assistance surfaces from the "interaction between the mediator and the learner, and is therefore highly sensitive to the learner's ZPD," and has little regard for the amount of time needed or for any preset outcomes (Poehner, 2008, p. 18).

I.1 Standardized Assessment versus Dynamic Assessment

Sternberg and Grigorenko (2002) discuss three main ways in which DA differs from SA. These criteria centre on development, feedback, and examiner involvement. Firstly, SA "taps more into a developed state, whereas dynamic testing taps more into a developing process" (Sternberg & Grigorenko, 2002, p. 28). Secondly, in SA, feedback during assessment is viewed as a hindrance to the testing process and, therefore, is avoided. In DA, however, feedback during assessment is encouraged and is often provided in varying forms (i.e. implicit to explicit). Lastly, in SA, the examiner tries not to be involved directly during the testing process itself. In DA, however, the examiner creates a setting of teaching and mediation in order to help the examinee improve in the process of testing. Furthermore, Lantolf and Thorne (2006) argue that the main component

that categorizes DA from SA is that DA creates a unity between instruction and assessment. This is further exemplified by Poehner (2008), who states, “DA overcomes the assessment-instruction dualism by unifying them according to the principle that mediated interaction is necessary to understand the range of an individual’s functioning but that this interaction simultaneously guides the further development of these abilities” (p. 24). DA integrates instruction and assessment into one process of learning and development, with the desired outcome of learner improvement during the assessment process (Poehner, 2008).

In the SLA field, there is a general agreement among researchers on the important role of corrective feedback in the process of second language acquisition (e.g. Sheen, 2011) and in the role of corrective feedback within a learner’s ZPD (e.g. Nassaji & Swain, 2000). Aljaafreh and Lantolf (1994) explain two criteria to which corrective feedback provided within the ZPD should adhere. Firstly, corrective feedback must be “graduated,” which means that feedback starts as implicit (e.g. Step 1, Figure 1) and gradually increases in specificity to be more explicit (e.g. Step 4, Figure 1) (p. 486). Secondly, feedback must be “contingent,” which means its provision must depend on when, or whether, the examinee requires it (p. 468). Moreover, corrective feedback affords learners with “dialogically negotiated assistance as they move from other-regulation towards self-regulation” (Lyster, Saito, & Sato, 2013, p. 9). In this way, corrective feedback and the ZPD are complimentary processes, which has implications for the use of corrective feedback in DA.

Aljaafreh and Lantolf (1994) conducted a descriptive study using this type of graduated and contingent corrective feedback working within the learner’s ZPD (Nassaji & Swain, 2000, p. 36). In this study, learners received one-on-one corrective feedback on written essays. This feedback was provided as a 12-step *regulatory scale*, which ranged from implicit to explicit. Following from this research, Nassaji and Swain (2000) conducted a similar study where they compared two methods of corrective feedback. This involved one method following the regulatory scale, which took the learner’s ZPD into consideration, and one using the scale randomly, therefore ignoring the learner’s ZPD. Therefore, following the regulatory scale put forth by Aljaafreh and Lantolf (1994), the feedback provided within the learner’s ZPD was collaborative and graduated, while the other feedback was random. Consequently, they found that the former was more effective (Nassaji & Swain, 2000).

I.2 Critiques of DA

There are several critiques of DA, which often stem from the way that traditional SA is viewed. First, proponents of SA claim that the focus DA places on learner development in the process of assessment risks the test’s “internal-consistency reliability” (i.e. whether learner performance is stable throughout the test) (Poehner, 2008, p. 71). This is a potential problem because the entity being assessed is altered and, therefore, can no longer be accurately determined (Poehner, 2008). For proponents of DA, however, this change in learner ability is evidence of a successful procedure precisely because the individual is learning (Poehner, 2008). This is one of the main areas in which SA and DA seem to be irreconcilable; SA finds problematic one of the core aims and outcomes of DA.

Further critiques are associated with the concepts of reliability and validity. Reliability refers to the extent to which a test-taker would receive the same results on a test over repeated instances (Huang, 2013). In terms of reliability, critics of DA are concerned that its interactionist nature makes reliability vulnerable because the provision of various mediations at different points in time may affect outcomes (Poehner, 2008). This critique is somewhat less significant in interventionist DA because certain levels of standardized procedures ensure the same assistance and feedback is given to each learner. Validity refers to the extent to which a test measures what it is seeking

to measure (Huang, 2013). Similar to the critique of reliability in DA, the same argument can be made for validity. Poehner (2008) explains that for DA, standard methods for ensuring validity are compromised by the assessment's aim of learner improvement. However, because development is a desirable outcome in DA, it is possible that the validity of DA should instead be understood as the degree to which it encourages development (Poehner, 2008). Because this study assumes that the goal of assessment is improving learners' English language proficiency, Poehner's argument is noteworthy.

In addition to concerns over reliability and validity, other criticism involve generalizability, which refers to the extent to which one can make predictions about learner performance in non-assessment contexts based on performance during assessment (Poehner, 2008). In SA, it is assumed that a learner's performance on a test should be able to accurately predict how they perform in non-assessment contexts. In DA, however, this emphasis on generalizability is less crucial because this dichotomy between assessment and non-assessment contexts is minimized by the merging of instruction and assessment (Poehner, 2008). Furthermore, commenting on DA's potential incompatibilities with traditional test criteria, Poehner (2008) argues that, "it should be clear that DA's incompatibility with more traditional frameworks does not invalidate it as an approach to assessment. Rather, their incommensurability simply points to the need for DA researchers to outline their own methods" (p. 73). In this way, DA's potential lack of standardized notions of reliability, validity, and generalizability does not necessarily invalidate the method, but rather indicates a need for a different outlook on these criteria.

I.3 Recent Major Studies in DA

In Lantolf and Poehner's (2014) most recent synthesis of DA in Applied Linguistics, they point out that the application of DA in the field has been sparse, and did not really begin until their own research in the early 2000s. However, there were a few studies that appear to have implemented a DA-like format without using the actual term dynamic assessment (e.g. Nassaji & Swain, 2000). That being said, research on DA in Applied Linguistics has been growing since these early studies. This section reviews some of the recent DA-related studies in the field of Applied Linguistics.

Poehner (2008) implemented a DA program focusing on the oral communication skills of six advanced French learners. Throughout his study, he found that learners' conceptual understanding of the verbal aspect in French developed uniquely through DA. Results also indicated that learners became more agentive (i.e. were independently active in improving their speaking skills) in their own learning, using various strategies to overcome obstacles. Overall, Poehner (2008) found that learners stretched beyond their current abilities to more complex tasks in DA.

Like Poehner, Antón (2009) also found positive results in DA. She conducted a study with five participants from an advanced Spanish language program, analyzing DA's role in the diagnostic assessment of speaking and writing levels. Her qualitative analysis indicated that DA allows for a deeper and fuller description of learners' actual and developing abilities. She argues that this allows programs to individualize and tailor the instruction to learners' specific needs.

In addition to Antón, Travers (2010) also used DA to investigate speaking levels. He compared IELTS™ speaking tests in SA and DA formats in a novel application of DA used to modify a standardized speaking test format. In his study, he sought to ascertain DA's potential advantage in regards to participant performance over independent speaking tests as well as examine the potential use of DA for learners from individualist versus collectivist cultural backgrounds. His findings from seven participants showed no difference in terms of mean test scores or gains over successive tests, but did show differences for grammar and lexical scores.

These three studies all had unique aims and outcomes which have greatly contributed to DA

research in the field of Applied Linguistics. However, there is still a need for research investigating quantitative test scores in DA. Moreover, further research into the interaction between language learner strategy use and performance using the DA method is needed. This is particularly pertinent when examining Canale and Swain’s (1980) prominent theory of communicative competence, which cites strategic competence as one of its main components (Huang, 2013). To my knowledge, the current research is the first study with this unique focus.

II. Language Learner Strategies

Language learner strategies or strategic behaviours are defined as the “conscious, goal-oriented thoughts and actions that learners use to regulate cognitive processes with the goal of improving language learning or language use” (Huang, 2013, p. 5). Language learner speaking strategies are often grouped into six major categories: approach, metacognitive, cognitive, communication, affective, and social (Huang, 2013). These six major categories of strategies and their definitions are outlined in Table 1.

Table 1: *Six Major Strategy Categories and Definitions (Huang, 2013, p.7)*

Strategy Category	Definition
Approach	Orienting oneself to the speaking task
Metacognitive	Examining the learning process in order to organize, plan, and evaluate efficient ways of learning
Cognitive	Manipulating the target language for understanding and producing language
Communication	Involving conscious plans for solving a linguistic problem in order to reach a communication goal
Affective	Involving self-talk or mental control over affect
Social	Interacting with others to improve language learning/use

Over the past four decades, the use of language learner strategies in the language learning context has been shown to positively affect language learning (Huang, 2012). Studies have shown a significant correlation between strategy use and success in language learning (e.g. Green & Oxford, 1995; Wong & Nunan, 2011). Huang (2010) contends that studies indicate that the use of metacognitive (e.g. Purpura, 1999) and cognitive (e.g. Oxford & Ehrman, 1995) strategies have a positive correlation with proficiency level, and in turn, this indicates the potential benefits of employing certain types of strategies. However, recent research supports the assumption that these strategies may not necessarily have a positive correlation with learner performance across all learners and situations, as certain specific strategies may have negative effects in testing or high-pressure contexts (e.g. Huang, 2013; Swain et al., 2009). Furthermore, learner strategies allow students to take more responsibility in their own learning and development (Wong & Nunan, 2011). Cohen (1998, as cited in Grenfell & Macaro, 2007) claims that language learning strategies can help learners shoulder more of the responsibility in their own learning, rather than relying solely on the teachers. In results from his survey of experts in the language learner strategy field, Cohen (2007) ascertained five purposes for language learning strategies: to enhance learning; to perform specified tasks; to solve specific problems; to make learning easier faster, and, more enjoyable; and, to compensate for a deficit in learning. It is evident that many experts agree that language learner strategies have a range of purposes that have the potential to enhance language learning.

The examination of strategy use is an important component of SLA research. As discussed, Canale and Swain’s (1980) influential framework of communicative competence includes strategic competence as one of its main components (Huang, 2013). Strategic competence can be defined as the, “learners’/speakers’ ability to use communication strategies to deal with communication

breakdowns” (Huang, 2013, p. 6). Furthermore, Bachman (1990) put forth that strategic competence was not only important for purely communicative interactions, but that it also served an executive function in all domains of language learning and use (Huang, 2013). In this way, the examination of strategy use in relation to assessment is crucial for SLA research. Moreover, Huang (2013) argues that even though there is recognition that strategies can potentially affect learner performance, there remains a gap in the research about the strategic component in speaking contexts, and about the specific interaction between strategic competence and second-language performance.

II.1 Critiques of Language Learner Strategies

Despite the research on the effectiveness of language learner strategies, there have been some criticisms in this area. Early criticisms of language learner strategies questioned the efficacy of the verbal expression of strategy use. For example, Seliger (1983, as cited in Grenfell & Macaro, 2007) postulated that researchers cannot assume that the verbalizations of learners correspond to internal mechanisms because strategies are so deeply rooted in the mind of the learner. However, White, Schramm, and Chamot (2007) outlined several methods in which the internal processes of strategy use are more apparent, including retrospective interviews, self-report questionnaires, reflection journals, and think-aloud protocols. For example, in her study on the use of oral reflection in facilitating oral production and strategy use, Huang (2012) found that gains over pre- and post-test scores could be indicative of oral reflection as an effective meditational tool to help learners progress in their use of metacognitive strategies. Moreover, White et al. (2007) discuss the efficacy of the successful elicitation of think-aloud protocols, which should include careful orientation, practice, and prompting, multiple language use, and “integration into an authentic action context” (p. 115).

For the purposes of this study, a think aloud, also called video-stimulated recall, refers to a method that involves videotaping a participant doing a task and then replaying the video for the participant as a prompt to elicit his or her thoughts on content that is in the scope of the research (Huang, 2014). Video-stimulated recall has the potential to “capture and investigate the dynamic nature of task performance” and, as such, one of its main purposes is “the potential to provide a wealth of information on the cognitive processes and strategic behaviours that participants engage in and deploy as they carry out a particular task or tasks across types or contexts” (Huang, 2014, p. 3). Several studies have pointed to the efficacy of using stimulated recall in the examination of language learner strategies (e.g. Sime, 2006; Macaro, 2006; Huang, 2014). For a detailed discussion of this method and its pros and cons, refer to Huang (2014).

III. Research Questions

The current research aims to address the following two research questions: 1) Is there a difference amongst students in the research sample between standardized and dynamic testing in terms of holistic (overall) and analytical (individual section) test scores? and 2) Do students in the research sample use different strategic behaviours in SA and DA contexts?

III. METHODS

I. Participants

Seven participants took part in this study. Participants, who speak Mandarin as a first language (L1), were recruited from the undergraduate student body of a mid-sized university in Western

Canada. I chose Mandarin-speaking students because there is a large population of Chinese students at this university. I chose only undergraduate students because I wanted to control for age and proficiency level to minimize individual learner variables. All participants were advanced learners of English. Table 2. presents participant demographics.

Table 2: *Participant Demographics*

Criteria	Category and Responses	
Age	Mean	19.3
	Range	18-23
Length of Residence	Mean	1 year
	Range	3 months-2 years
Gender	Male	4 (57.1%)
	Female	3 (42.9%)
First-Language	Chinese	
Test Taking Background	IELTS (6)	Score Range: 5.5-6.5
	TOEFL (1)	Score: 86

Before beginning participant recruitment, I received ethics approval from The University of Victoria Human Research Ethics Board (Protocol Number 14-360). I recruited participants through several methods including email, posters, and Facebook postings. Each recruitment method outlined the purpose of the study and the steps involved in participating, as per the university's ethical guidelines.

II. Instruments

II.1 Background Questionnaire

All participants completed a background questionnaire. This questionnaire, adapted from Huang (2012), included questions on gender, age, degree program, possible linguistic qualifications, first contact with English, length of learning English, length of residence in Canada, English usage per day, other languages spoken, and standardized test-taking background.

II.2 IELTS™ Speaking Test

I used the International English Language Testing System (IELTS™) Speaking test to conduct the study. IELTS™ is a widely accepted test of English language proficiency worldwide and is used by 9,000 organizations in over 140 countries (IELTS, 2013). The IELTS™ Speaking test consists of three sections, one of which was used in this study. I adapted the test to consist of only the first task, which involves general questions about familiar topics, such as home, family, work, studies, or interests (IELTS, 2013). The other sections were not used because the second section consists of the examinee speaking about a topic for several minutes uninterrupted; therefore, it would not have been feasible to adapt this portion to a DA format. Likewise, I could not use the third section because many questions in this section depend on the answers that learners provide in the second

section. The first section usually involves only four to five questions and lasts approximately four to five minutes; therefore, the duration and scope were not enough to adequately measure a participant's speaking ability. In consultation with a certified examiner and a language-testing specialist, I extended the first task. In order to do this, I incorporated two task-ones and added two, randomly selected longer opinion questions from task three to each test. Therefore, each test was divided into two topics, with each topic having four shorter questions about basic information (from task 1) and one opinion question that required a longer answer (from task 3). These tests were used for both the standardized and the dynamic version of the test.

II.3 Regulatory Scale for DA

For the DA procedure, I used the regulatory scale as adapted from Travers (2010) and Aljaafreh and Lantolf (1994). As mentioned, this corrective feedback scale ranges from implicit (e.g. Step 1, Figure 1) to explicit (e.g. Step 4, Figure 1). Travers's scale included four steps, but I modified the scale to follow five steps because the provision of the extra step (i.e. Step 4) would offer the participants an additional opportunity to self-correct. Additionally, I decided to provide a recast (rather than the correct utterance) as the final step because I did not want the correction to affect participant performance on the remainder of the test. The scale used in the present study is outlined in Figure 1.

1. Examiner indicates that something may be wrong in a speaking turn ("Sorry").
2. Examiner narrows down the location of the error (e.g. examiner repeats the specific speaking turn that contained the error).
3. Examiner indicates the nature of the error, but does not identify the error (e.g. "There was something wrong with the tense marking there").
4. Examiner provides clues to help the learner arrive at the correct form (e.g. "It is not really past but something that is still going on").
5. Examiner provides a recast (no matter if utterance is correct or incorrect).

Figure 1: *Regulatory Scale*

III. Data Collection Procedures

III.1 Outline of Procedure

Data collection consisted of two sessions over two days. Each session included one test (standardized or dynamic) followed by video-stimulated recall. The first session began with informed participant consent followed by administration of the background questionnaire. I then administered the first test. Each participant took one standardized test and one dynamic test, and I acted as the examiner for all tests (i.e. standardized and dynamic). I randomized the test order, so that some participants took the standardized test the first day and the dynamic test the second, while some did the opposite. Each participant took two different tests with different questions, to minimize potential practice effects regarding topic. All the tests were video recorded in order to facilitate the video-stimulated recall, which immediately followed the testing. All video-stimulated recall sessions were conducted by a research assistant, who is a graduate student in an Applied

Linguistics program. At the end of the recall session, the research assistant asked the participant a few more follow-up questions concerning whether he or she treated it as a real test and what he or she liked and disliked about the test itself. The second session started with a verification of ongoing participant consent. Then, I administered the second test, which consisted of the testing method not used in the first session. This was again followed by video-stimulated recall and the same follow-up questions as the first session, with the exception of one additional question, which asked the participant to compare the two tests.

III.2 Standardized Testing Procedure

I conducted the standardized IELTS™ Speaking tests following IELTS™ test administration guidelines. I asked the questions and did not give any feedback other than providing neutral responses (“okay”), answering clarification requests, such as repeating the question or explaining the meaning of a word (S. Abrar-ul-Hassan, personal communication, Nov. 14, 2015), and encouraging the participant to elaborate with responses such as “Why?” or “Can you tell me more about that?” in cases where he or she did not provide enough information.

III.3 Dynamic Testing Procedure

I conducted the dynamic tests in a more interactive manner. This study followed a more interventionist type of DA due to the nature of the research and the need for quantifiable results, following previous studies (e.g. Travers, 2010). Therefore, the dynamic method involved providing graded levels of corrective feedback, following the regulatory scale (Figure 1), when the participant made a grammatical mistake. I followed these steps until the participant self-corrected or, in the case of no self-correction, until the last step (i.e. recast). I chose to focus solely on grammatical errors and of those, on common errors made by learners with a Chinese L1. This includes areas such as inflection of gender, number, and case, subject-verb agreement, verb tense, progressive aspect, absence of “be” before predicative adjectives, prepositions, articles, and count/non-count nouns (Swan & Smith, 2001). I did this in order to narrow down the amount of errors for research purposes.

III.4 Video-Stimulated Recall

The video-stimulated recall involved the research assistant and the participant watching the video of the testing process just conducted. Throughout this time, the research assistant paused the video periodically when the participant did something of interest (e.g. pausing, self-correcting, attending to examiner feedback) and asked the participant to comment on what he or she was thinking at that moment. This included questions such as, “What were you thinking before the test?” “What were you thinking here?” “I noticed you did X here. What were you thinking then?” and “When the examiner said that, what were you thinking?” The participant also had the option to pause the video and comment at any point. The research assistant is a Mandarin-as-a-first-language speaker; therefore, participants were given the option to speak in or be spoken to in their L1 at any time throughout the stimulated recall.

III.5 Scoring with IELTS™ Descriptors

The first stage in my data analysis involved scoring the tests with the IELTS™ Speaking Descriptors, which are measured on a 9-point scale (IELTS, 2014). These descriptors encompass four different analytical sections, including fluency and coherence, lexical resource, grammatical range

and accuracy, and pronunciation. I scored both the standardized and dynamic tests, analytically and holistically, using these descriptors. Additionally, the research assistant independently scored the tests using these descriptors, in order to ensure inter-rater reliability. We then compared our scores, and in cases where they showed a difference of more than 0.5 points, we discussed the score until we agreed on a score within 0.5 of each other. In order to get the final scores, I then averaged both of our scores for each analytical component, so as to have one score for each component, and averaged these to calculate the holistic scores.

III.6 Transcribing and Coding Strategies

In order to analyze the data from the video-stimulated recall sessions, I first transcribed all the video clips. The research assistant translated and transcribed the portions that were in Mandarin. I then coded all the transcript data for both the reported and observed strategies. After I had coded all the strategies, I crosschecked one participant transcript with my supervisor, in order to have inter-rater reliability. I also recoded the data from 43% of participants, without referring back to the original, and then crosschecked these codes with my initial ones. My intra-rater reliability was 82%. After coding, I grouped these individual strategies into major categories including approach, metacognitive, cognitive, communication, affect, and social. I then calculated the percentage of use of each major strategy category (including both reported and observed strategies) for each method of assessment. I also counted the number of occurrences for each individual strategy and calculated the percentage of use with regard to the major strategy categories as well as to the total number of codes within that method of assessment. Lastly, I determined the top scoring and low scoring participants for each assessment type and analyzed each of these participants for strategy use.

IV. RESULTS AND DISCUSSION

Research Question 1: *Is there a difference amongst students in the research sample between standardized and dynamic testing in terms of holistic and analytical test scores?*

Out of seven participants, six reported treating the simulated exam as a real test. Holistically, there was very little difference (i.e. Table 1, column 2 & 8) in test scores between standardized and dynamic tests. Analytically, however, there were several differences. Scores for fluency and coherence showed variation, with 71% of participants scoring higher in SA. Scores for lexical resource also showed variation, with 57% of participants scoring higher in DA. Similarly, in the category of grammatical range and accuracy, 57% of participants scored higher in DA. Scores for pronunciation did not differ much between the two methods. Table 2 summarizes test scores.

The difference in fluency scores could potentially be due to the method of corrective feedback used in the DA procedure. When participants made a grammatical error, I would disrupt their flow of speech in order to help them self-correct. Therefore, it is possible that this mediation interrupted their fluency. The format of the dynamic test could also be a factor in the variation of lexical and grammatical scores. My sole focus on grammatical errors could potentially play a role in participants' learning potential in this area. This is in congruence with results from Travers (2010). Furthermore, it is possible that this focus on grammatical errors positively contributed to lexical resource scores, as the criteria for this category are associated with paraphrasing and the use of idiomatic phrases, both of which can be influenced by grammatical ability. Amongst the sample participants, the results indicate that DA is better able than SA to facilitate certain language abilities (i.e. grammatical and lexical); therefore, assuming the goal of assessment is improving language proficiency, DA is potentially better able to achieve this goal amongst the

Table 3: Participant Test Scores

P	Standardized					Dynamic					
	Hol	Flu	Lex	Gram	Pron	P	Hol	Flu	Lex	Gram	Pron
1	6.0	5.8	6.0	6.3	6.0	7	5.9	5.8	5.8	6.0	6.0
3	5.9	5.8	5.5	5.8	6.3	3	5.8	5.3	5.8	6.0	6.0
7	5.9	6.0	5.8	5.8	6.0	2	5.7	5.8	6.0	5.0	5.8
6	5.8	6.0	5.8	5.5	6.0	4	5.7	5.0	5.5	6.0	6.3
2	5.7	5.5	5.5	5.8	6.0	6	5.7	5.8	5.5	5.8	5.8
5	5.7	5.5	5.3	5.8	6.0	5	5.6	5.5	5.5	5.3	6.0
4	5.6	5.3	5.3	5.5	6.3	1	5.5	5.3	5.3	5.5	6.0
M	5.8	5.7	5.6	5.8	6.1	M	5.7	5.5	5.6	5.7	6.0

Note. P= participant; Hol= holistic; Flu= fluency and coherence; Lex= lexical resource; Gra= grammatical range and accuracy; Pro= pronunciation; M= mean.

sample. However, this may not be the case if the goal of assessment is considered differently. It is also important to note that this type of descriptive statistical analysis is meant only to generalize amongst the sample participants, and cannot be used to generalize real differences between the two methods of assessment in the population.

When asked about their test preference, three participants reported preferring the standardized test, three reported preferring the dynamic test, and one had no preference. Of the participants who preferred the standardized test, two said it was because stopping for corrective exchanges in the dynamic test made them more nervous. It is possible that due to the small scale of this research, with participants only taking one test of each type, there was not sufficient time for participants to become comfortable with the DA testing process. This is supported by Travers (2010), who examined gains over three successive tests, and found that out of the participants who reported being nervous or uncomfortable with the corrective interruptions in the DA format in early tests, all reported being comfortable with these corrections by the third test. Additional reasons for test preference included order of test taking, question preference, and interest in the topic (i.e. If a certain test topic was more interesting to the learner, they sometimes cited preferring that test).

Research Question 2: Do students in the research sample use different strategic behaviours in standardized and dynamic assessment contexts?

In total, I recorded and observed that participants used 33 different individual strategies with 270 occurrences across the two methods of assessment. Within these strategies, four were approach strategies, twelve were metacognitive strategies, nine were cognitive strategies, five were communication strategies, one was an affective strategy, and two were social strategies. Figure 2 illustrates strategy use in SA and DA.

Overall, approach and communication strategy use was similar across both methods of assessment. Instances of metacognitive strategy use were greater in DA than in SA, and for both methods, this was the type of strategy used the most in comparison with other strategy categories. Although the proportion of metacognitive strategy use was the same (48%) for both DA and SA, there were more individual instances of use for DA (78 versus 52). Cognitive strategies were also greater in DA. Previous studies have shown a positive correlation between metacognitive (e.g. Nakatani, 2005) and cognitive (e.g. Oxford & Ehrman, 1995) strategies and language proficiency (e.g. Huang, 2010). Anderson (2005) proposes that metacognitive strategy use is more effective because “once a learner understands how to regulate his or her own learning through the use

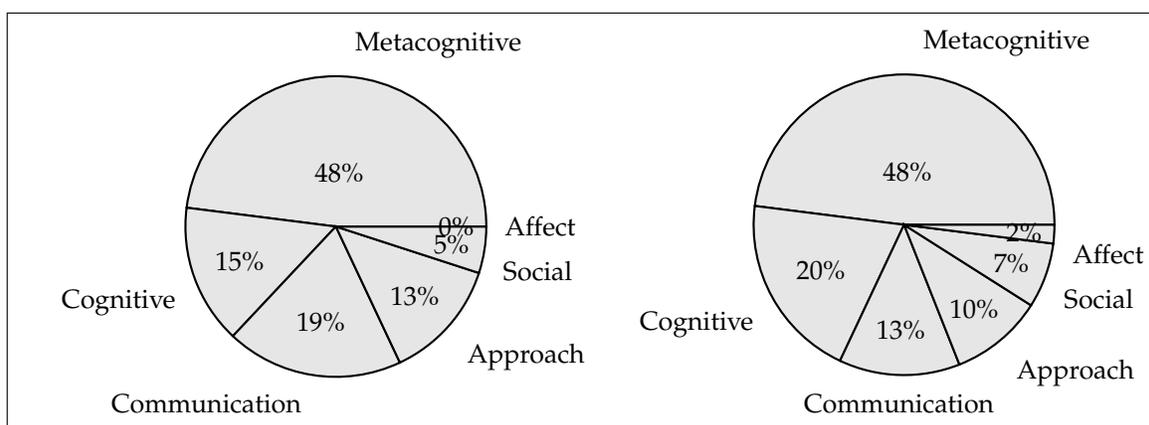


Figure 3: Reported and Observed Strategy Categories for SA (n= 107) (left) and DA (n= 163) (right)

of strategies, language acquisition should proceed at a faster rate” (p. 766). Affective strategy use was also greater in DA in the study sample; however, the only affective strategy recorded was expressing affect in response to correction, which relied on correction only provided in the DA format. Lastly, social strategies were higher in DA, as is expected from the interactive format of the DA method. Differences in individual strategy use revealed that SA showed greater instances of evaluating test performance, generating words, seeking clarification, elaborating, and self-assessment, while DA showed greater instances of evaluating oral production, attending to feedback, formulating ideas, and L1 translation.

In addition to an examination of individual strategy use, I also determined top scorers in each group to be any participants who scored at or above the mean score; this corresponded to four participants for SA and five participants for DA. The only common top scorer strategy used in both SA and DA was self-correction. The different strategy use between top scorers for the two methods of assessment may be indicative of context-/method-specific strategy use, rather than based on proficiency level, as three of the participants were top scorers in both SA and DA and showed little commonality. These results are supported by research on learner strategy use and context, which indicates a variation in strategy use across contexts (e.g. Huang, 2013; Swain et al., 2009). Specifically, in her study of the use of language learner strategies in testing and non-testing contexts, Huang (2013) found differences in strategy use between these two contexts as well as between different IELTS™ tasks, but did not find significant differences in strategy use in regards to proficiency level. This is further supported by White et al. (2007) who argue, “strategy use is not a fixed attribute of individuals, but changes according to the task, the learning conditions, and the available time” (p. 93).

Low scorers, those who scored below the mean score, also demonstrated some differences. In DA, the two lowest scorers showed the highest individual use of self-correction and evaluating oral production, which indicates a limited repertoire of strategies. Studies have shown that successful language learners have a more extensive range of strategies and utilize a variety of individual strategies in the process of language learning (Anderson, 2005). However, this does not necessarily mean that higher strategy use always correlates to better performance. Anderson (2005) claims that it is important that learners apply strategies in an effective manner, and that this is more significant than number of strategies or particular strategy use. Similarly, Huang (2010) argues that “what matters for individual learners is not accumulating or using a wide variety of strategies, but managing a repertoire of strategies” (p. 19).

I. Pedagogical Recommendations

The results from this study brought to light several noteworthy pedagogical recommendations:

- a. **Familiarize learners and teachers with the DA method.** As individual participant's comfort with the corrective feedback exchanges varied within the sample, it is important to ensure that the learner and teacher understand and are familiar with the DA method in order for the assessment to be productive for both. If the teacher was familiar with the DA procedure, then he or she would be better able to assist and scaffold the learner within his or her ZPD.
- b. **Tailor DA to learner proficiency level.** The participants in this study were all advanced learners of English, and, as such, this could have been a factor in their response to corrective feedback in regards to affect and receptiveness. It is possible that for advanced learners of English, DA should be less interruptive in order to avoid disrupting fluency or idea coherence and to increase learner receptiveness.
- c. **Focus DA on linguistic features where learners need to develop.** DA can potentially facilitate grammatical and lexical competence of learners, as indicated by the differences in participants' grammatical range and accuracy and lexical resource scores. The foci of mediation were imperative to outcomes in participants' DA scores.
- d. **Attend to the strategies used by learners in DA.** It is possible that the DA method potentially promoted the use of metacognitive and cognitive strategies amongst participants. As discussed, research shows that use of these strategies tends to correlate positively with learner oral production in the learning context. It is important to attend to strategies because awareness, for the learner as well as for the examiner or instructor, is essential to the successful use of strategies in language learning.
- e. **Promote learner agency through DA.** It is possible that DA promoted learner agency amongst the participants in the sample. This is indicated by the much greater use of strategies such as evaluating oral production and attending to feedback in DA. This is in congruence with results from Poehner's (2008) study, which revealed learners exhibiting agency through the use of various strategies during DA. These particular strategies promoted learner agency because they allowed the learner to evaluate inwardly and independently find a solution to a language problem. Learner agency is a good thing for both learners and instructors because it allows the learners to develop independent language problem solving skills that they can use on their own without relying on the instructor, and consequently, this takes some of the responsibility off of the instructor (Cohen, 1998, as cited in Grenfell & Macaro, 2007).

II. Limitations and Future Research Directions

The results from this study should be considered in light of the following three research limitations. First, at times during the dynamic test, participants did not realize that they had made a mistake or did not understand my attempt to help them self-correct. Moreover, because I chose to follow a standardized scale of corrective feedback in order to enhance reliability for research purposes, I could not answer when a participant asked a question that veered from the responses based on the corrective scale. Second, because the study consisted of only seven advanced learners, further research should be conducted involving a larger sample size with learners of different proficiency levels in order to decipher other potential differences in areas such as test scores (holistic and analytical) and strategy use through statistical analysis. Third, because the test format included only the first section of the IELTSTM Speaking test, further research should examine the role of DA

in performing a full-scale standardized test with a focus on other linguistic features. Additionally, there are several pedagogical limitations. It is possible that including recast in the regulatory scale (rather than the correct answer) may have undermined participant learning. In addition, the DA method is quite time-consuming and, as such, the method may not be feasible for large class sizes. Rather, it is more conducive for workshops or English-for-specific-purposes, with instruction tailored to the needs of individual learners.

V. CONCLUSION

This exploratory research has examined the differences between SA and DA in terms of speaking test scores and language learner strategies. It has highlighted differences in terms of fluency, grammar, and lexical resource. Test scores were higher for fluency in SA, while scores were higher for grammar and lexical resource in DA. This points to the potential of DA as a mediating tool to facilitate the development of grammatical and lexical abilities amongst the participants. Moreover, strategy use was greater in DA, particularly for metacognitive and cognitive strategies, which tend to correlate positively with learner proficiency and performance in the learning context. Empirically, the current study has contributed to the research on DA in the fields of SLA and Applied Linguistics. Practically, this study can inform instructors about the best method of classroom-based assessment for improving English language proficiency that goes beyond standardized testing, in order to facilitate the potential for learners to improve their English-speaking skills.

REFERENCES

- Anderson, N.J. (2005). L2 strategy research. In E. Hinkel (Ed.), *Handbook of research in second language teaching and learning* (pp. 757-772). Mahwah, NJ: Lawrence Erlbaum Associates. <http://dx.doi.org/10.4324/9781410612700>
- Antón, M. (2009). Dynamic assessment of advanced second language learners. *Foreign Language Annals*, 42(3), 576-598. <http://dx.doi.org/10.1111/j.1944-9720.2009.01030.x>
- Bachman, L. F. (1990). *Fundamental considerations in language testing*. Oxford, UK: Oxford University Press.
- Budoff, M. (1987). The validity of learning potential assessment. In C.S. Lidz (Ed.), *Dynamic assessment: An interactive approach to evaluating learning potential* (pp. 52-81). New York, NY: Guilford.
- Canale, M., & Swain, M. (1980). Theoretical bases of communicative approaches to second language teaching and testing. *Applied Linguistics*, 1, 1-47. <http://dx.doi.org/10.1093/applin/1.1.1>
- Cohen, A.D. (2007). Coming to terms with language learner strategies: Surveying the experts. In A.D. Cohen & E. Macaro (Eds.), *Language learner strategies* (pp. 29-45). Oxford, UK: Oxford University Press.
- Feuerstein, R., Rand, Y., & Rynders, J.E. (1988). *Don't accept me as I am. Helping retarded performers excel*. New York, NY: Plenum.

- Green, J.M., & Oxford, R.L. (1995). A closer look at learning strategies, L2 proficiency, and gender. *TESOL Quarterly*, 29(2), 261–297. <http://dx.doi.org/10.2307/3587625>
- Grenfell, M., & Macaro, E. (2007). Claims and critiques. In A.D. Cohen & E. Macaro (Eds.), *Language learner strategies* (pp. 9-28). Oxford, UK: Oxford University Press.
- Huang, L.-S. (2010, Spring). Key concepts and theories in TEAL: Language learner strategies. *TEAL News: The Association of B.C. Teachers of English as an Additional Language*, (pp. 18-20). Retrieved from http://www.bctéal.org/wpcontent/uploads/2011/08/BCTEALNews_Spring_2010.pdf
- Huang, L.-S. (2012). Use of oral reflection in facilitating graduate EAL students' oral language production and strategy use: An empirical action research. *International Journal for the Scholarship of Teaching and Learning (IJ-SoTL)*, 6(2), 1-22.
- Huang, L.-S. (2013). Cognitive processes involved in performing the IELTS Speaking test: Respondents' strategic behaviours in simulated testing and non-testing contexts, (pp. 51). *IELTS Research Report Series*. Retrieved from http://www.ielts.org/pdf/Huang_RR_Online_2013.pdf
- Huang, L.-S. (2014). Key concepts and theories in TEAL: Cognitive validity. *SHARE: TESL Canada's eMagazine for ESL Teachers*.
- Huang, L.-S. (2014). Video-stimulated verbal recall: A method for researching cognitive processes and strategic behaviours. In *SAGE Research Methods Cases* (pp. 1-22). London: SAGE Publications.
- IELTS Web site. (2013). Retrieved March 20, 2015, from www.ielts.org
- Lantolf, J.P., & Poehner, M.E. (2014). *Sociocultural Theory and the pedagogical imperative in L2 education: Vygotskian praxis and the research/practice divide*. New York, NY: Routledge. <http://dx.doi.org/10.4324/9780203813850>
- Lantolf, J.P., & Thorne, S.L. (2006). *Sociocultural Theory and the genesis of second language development*. Oxford, UK: Oxford University Press.
- Lyster, R., Saito, K., & Sato, M. (2013). Oral corrective feedback in second language classrooms. *Language Teaching*, 46(1), 1-40. <http://dx.doi.org/10.1017/s0261444812000365>
- Macaro, E. (2006). Strategies for language learning and for language use: Revising the theoretical framework. *Modern Language Journal*, 90, 320-337. <http://dx.doi.org/10.1111/j.1540-4781.2006.00425.x>
- Nakatani, Y. (2005). The effects of awareness-raising training on oral communication strategy use. *The Modern Language Journal*, 89(1), 76-91. <http://dx.doi.org/10.1111/j.0026-7902.2005.00266.x>
- Nassaji, H., & Swain, M. (2000). A Vygotskian perspective on corrective feedback in L2: The effect of random versus negotiated help on the learning of English articles. *Language Awareness*, 9(1), 34-51. <http://dx.doi.org/10.1080/09658410008667135>

- Oxford R.L., & Ehrman, M. E. (1995). Adults' language learning strategies in an intensive foreign language program in the United States. *System*, 23(3), 359-386.
[http://dx.doi.org/10.1016/0346-251x\(95\)00023-d](http://dx.doi.org/10.1016/0346-251x(95)00023-d)
- Poehner, M. E. (2008). *Dynamic assessment: A Vygotskian approach to understanding and promoting L2 development*. Berlin, Germany: Springer.
- Poehner, M. E., & Lantolf, J. P. (2013). Bringing the ZPD into the equation: Capturing L2 development during computerized dynamic assessment (C-DA). *Language Teaching Research*, 17(3), 323-342. <http://dx.doi.org/10.1177/1362168813482935>
- Purpura, J. E. (1999). *Learner strategy use and performance and language tests: A structural equation modeling approach*. Cambridge, UK: Cambridge University Press.
- Sheen, Y. (2011). *Corrective feedback, individual differences and second language learning*. New York, NY: Springer.
- Sime, D. (2006). What do learners make of teacher's gestures in the language classroom? *International Review of Applied Linguistics in Language Teaching*, 44, 211-230.
<http://dx.doi.org/10.1515/iral.2006.009>
- Sternberg, R. J., & Grigorenko, E. L. (2002). *Dynamic testing: The nature and measurement of learning potential*. Cambridge, UK: Cambridge University Press.
- Swain, M., Huang, L.-S., Barkaoui, K., Brooks, L., & Lapkin, S. (2009). The speaking section of the TOEFL iBT™ (SSTiBT): Test-takers' reported strategic behaviors. *TOEFL iBT™ Research Series No. TOEFLiBT-10*. Princeton, NJ: Educational Testing Service.
- Swan, M., & Smith, B. (2001). *Learner English: A teacher's guide to interference and other problems*. Cambridge, UK: Cambridge University Press. <http://dx.doi.org/10.1017/cbo9780511667121>
- Travers, N. (2010). *Relating learner culture to performance on English speaking tests with interactive and non-interactive formats*. Retrieved from UVicSpace: Electronic Theses and Dissertations.
- Vygotsky, L.S. (1978). *Mind in society: The development of higher psychological processes*. Cambridge, MA: Harvard University Press.
- White, C., Schramm, K., & Chamot, A.U. (2007). Research methods in strategy research: Re-examining the toolbox. In A.D. Cohen & E. Macaro (Eds.), *Language learner strategies* (pp. 93-116). Oxford, UK: Oxford University Press.
- Wong, L.L.C., & Nunan, D. (2011). The learning styles and strategies of effective language learners. *System*, 39(2), 144-163. <http://dx.doi.org/10.1016/j.system.2011.05.004>