


2013

The Effect of Dynamic Assessment on Adult Learners of Arabic: A Mixed-Method Study at the Defense Language Institute Foreign Language Center

Mohsen Mahmoud Fahmy

University of San Francisco, mohsenfahmy730@gmail.com

Follow this and additional works at: <https://repository.usfca.edu/diss>

 Part of the [Adult and Continuing Education and Teaching Commons](#), and the [Linguistics Commons](#)

Recommended Citation

Fahmy, Mohsen Mahmoud, "The Effect of Dynamic Assessment on Adult Learners of Arabic: A Mixed-Method Study at the Defense Language Institute Foreign Language Center" (2013). *Doctoral Dissertations*. 85.
<https://repository.usfca.edu/diss/85>

This Dissertation is brought to you for free and open access by the Theses, Dissertations, Capstones and Projects at USF Scholarship: a digital repository @ Gleeson Library | Geschke Center. It has been accepted for inclusion in Doctoral Dissertations by an authorized administrator of USF Scholarship: a digital repository @ Gleeson Library | Geschke Center. For more information, please contact repository@usfca.edu.

The University of San Francisco

THE EFFECT OF DYNAMIC ASSESSMENT ON ADULT LEARNERS OF ARABIC:
A MIXED-METHOD STUDY AT THE DEFENSE LANGUAGE INSTITUTE
FOREIGN LANGUAGE CENTER

A Dissertation Presented
to
The Faculty of the School of Education
International and Multicultural Education Department

In Partial Fulfillment
of the Requirements for the Degree
Doctor of Education

by
Mohsen M. Fahmy
San Francisco, CA
December 2013

THE UNIVERSITY OF SAN FRANCISCO

Dissertation Abstract

The Effect of Dynamic Assessment on Adult Learners of Arabic: A Mixed-Method Study
at the Defense Language Institute Foreign Language Center

Dynamic assessment (DA) is based on Vygotsky's (1978) sociocultural theory and his Zone of Proximal Development (ZPD). ZPD is the range of abilities bordered by the learner's assisted and independent performances. Previous studies showed promising results for DA in tutoring settings. However, they did not use proficiency-based rubrics to measure students' progress and did not mention the method of using DA practically in classrooms. The literature showed that task-based language instruction (TBLI) is effective in adult classrooms. This study combined DA with TBLI to answer four questions. What is the change in the structural control of Arabic speaking based on DA/TBLI instruction? How do Oral Proficiency Interview (OPI) without DA assistance and OPI with DA assistance compare relative to the evaluation of Arabic speaking? How do the experiences and perceptions of DA/TBLI instruction compare between teacher-researcher and OPI testers? What are the student perceptions of the DA process? The study was conducted in three phases to answer its questions: pre-DA, DA, and post-DA. In the pre-DA phase, 12 volunteers from the Defense Language Institute Foreign Language Center went through unofficial Oral Proficiency Interviews (OPI), intellectual style survey, biographical background questionnaire, and interventionist-DA interviews. During the DA phase, the teacher-researcher used DA/TBLI instruction and Interagency Language Roundtable-based (ILR) rubrics to promote learning and to diagnose students' needs daily. These

lessons were observed by certified OPI testers. In the post-DA phase, the six selected participants were reevaluated by OPIs and interventionist-DA interviews. Students and observers were interviewed, but only students responded to a survey. The results of comparing the different evaluations conducted in both the pre- and post-DA phase showed that the structural control of Arabic improved for all participants. There is a parallel coefficient of 1.0 between the OPI with and without DA assistance for evaluating the participants' speaking proficiency. DA/TBLI instruction was practical and successful in making a difference for the participants' learning process. It reflected the success of the ILR-based rubrics in diagnosing accurately the students' inabilities whether in the interventionist-DA interviews or in the daily interactionist DA. The OPI without DA assistance cannot provide accurate diagnostic feedback in details.

This dissertation, written under the direction of the candidate's dissertation committee and approved by the members of the committee, has been presented to and accepted by the Faculty of the School of Education in partial fulfillment of the requirements for the degree of Doctor of Education. The content and research methodologies presented in this work represent the work of the candidate alone.

Fahmy, Mohsen

Candidate

12/9/2013

Date

Dissertation Committee

Cary, Stephen

Chairperson

12/9/2013

Busk, Patricia

12/9/2013

Taylor, Betty

12/9/2013

ACKNOWLEDGEMENTS

My path through finishing this dissertation had been a life-long journey from Egypt to the United States of America. Many great people provided me generously with indispensable support and encouragement. I consider myself the luckiest to attend the doctoral program at the University of San Francisco (USF) with its inclusive culture and challenging academic environment. I would not have been able to finish this dissertation without the informed guidance of my dissertation committee. Without Dr. Betty Tylor's ushering and eye-opening guidance, Dr. Patricia Busk's detailed and meticulous technical expertise and introducing me to performance-based assessment, and Dr. Cary's tireless and patient assistance, this dissertation would not have ended even close to its final quality. Dr. Cary provided me with numerous suggestions, insightful knowledge, and excellent editing that were priceless in fine-tuning this study.

My arrival to the University of San Francisco would not be possible without Dr. Jean Turner's exemplary dedication to her students. Her intimate understanding of international education and teaching in a multicultural environment not only enabled me to become an informed member of this field, but also inspired me to attempt following her admirable commitment to excellence. I owe her and the other esteemed professors at Monterey Institute of International Studies (MIIS) a life-long debt of gratitude for transferring to me their inspirational passion for the field of foreign-language teaching and loyalty to students. The profound and abreast knowledge of Dr. Kathleen Bailey and late Dr. Leo van Lier fueled my academic career of teaching and learning.

Being with exuberant educators such as the previously mentioned professors would not be accessible without the support of the monumental edifice of language

learning in the world, the Defense Language Institute Foreign Language Center (DLIFLC). The institute's commitment to language learning not only availed me great education at MIIS and USF, but it also provided me with diverse, multicultural, and experiential-learning venues for most disciplines in the field of foreign language education and research. The institute gave me an invaluable opportunity to learn firsthand from many authorities in the field of language learning and research such as Dr. Pat Boylan, Mr. Jim Child, Dr. Ray Clifford, Dr. Pardee Lowe, Dr. Madeline Ehrman, Dr. Gordon Jackson, Dr. John Lett, and Dr. Rebecca Oxford. I am grateful also for the relentless support of all the current and former managers and supervisors at DLIFLC for facilitating my academic pursuit and conducting this study. I am humbled also by the collegial assistance of DLIFLC's 22 certified testers and the generous volunteering of the institute's 10 students; their contributions were pivotal to the success of this study.

The emotional sustenance of my family and friends in Egypt and USA will remain in my heart and mind for as long as I shall live. I learned from the Egyptian Armed Forces that leadership is a selfless service for a bigger cause, esprit de corps, and task accomplishment. The unconditional love I enjoyed growing up in my family has bolstered my self-confidence to reach many formidable goals. My precious wife, Lisa Fahmy, provided me with her gracious care and encouragement, which maintained my morale and motivation through the ups and downs of dissertation writing. She did everything possible to help me focus on critical thinking and writing. In return, I dedicate this work to her as a token of my heartfelt appreciation and to my children, Janna and Adam Fahmy, to symbolize the importance of learning and my unconditional love to them too.

TABLE OF CONTENTS

	Page
ABSTRACT.....	ii
SIGNATURE PAGE.....	iv
ACKNOWLEDGEMENTS.....	v
TABLE OF CONTENT.....	vii
LIST OF TABLES.....	x
LIST OF FIGURES.....	xi
CHAPTER I RESEARCH PROBLEM	1
Statement of the Problem.....	2
Background and Need.....	5
Purpose of Study.....	8
Research Questions.....	10
Theoretical Rationale.....	11
Definition of Terms.....	16
Significance of Study.....	18
CHAPTER II LITERATURE REVIEW.....	20
Dynamic Assessment.....	20
<i>Important Concepts of Dynamic Assessment</i>	22
<i>ZPD and Its Use in Dynamic Assessment</i>	23
<i>DA's Interventionist Model</i>	25
<i>DA's Interactionist Model</i>	26
<i>Transfer of Learning</i>	28
<i>Previous DA Studies</i>	30
Task-Based Language Instruction.....	47
<i>Performance-Based Assessment</i>	50
<i>Previous TBLI Studies</i>	53
Summary.....	58
CHAPTER III METHODOLOGY	62
Research Design	64
Participants.....	68
Protection of Human Subjects.....	73

TABLE OF CONTENTS CONTINUED

	Page
Instruction.....	74
Instrumentation.....	77
<i>The Dynamic Assessment Rubrics Form</i>	79
<i>Validity and Reliability</i>	80
<i>The Validation Process of Rubrics</i>	81
<i>Guiding Questions for the Observers</i>	84
<i>Guiding Questions for Interviewing Students</i>	84
<i>Guiding Questions for the Teacher Journal</i>	85
<i>The Ten 5-point Scales</i>	86
Use of Assessment.....	86
<i>The Oral Proficiency Interview</i>	88
<i>OPI Structure</i>	92
<i>Tasks and Their Coverage</i>	93
<i>Reliability</i>	95
<i>Training Raters</i>	96
<i>Validity</i>	97
<i>Sensitivity to Instruction</i>	99
<i>Consequential Validity</i>	100
<i>Fairness and Equity</i>	100
<i>Text Typology</i>	102
The Teacher-Researcher.....	104
Research Questions.....	105
Data Analysis.....	106
CHAPTER IV RESULTS.....	108
Overview.....	108
Design Overview.....	108
Research Question 1.....	109
<i>Basem</i>	113
<i>Hazem</i>	115
<i>Ibrahim</i>	117
<i>Jamal</i>	119
<i>Salwa</i>	120
<i>Ramzy</i>	122
Research Question 2.....	125
Research Question 3.....	131
Research Question 4.....	135
Summary.....	139

TABLE OF CONTENTS CONTINUED

	Page
CHAPTER V SUMMARY, DISCUSSION, CONCLUSIONS, IMPLICATIONS.....	142
Summary of Study.....	142
Discussion.....	145
<i>Question 1</i>	145
<i>Question 2</i>	150
<i>Question 3</i>	154
<i>Question 4</i>	157
Conclusion.....	162
Recommendations.....	165
<i>Practices</i>	165
<i>Future Research</i>	166
Limitations of Study.....	169
REFERENCES	171
APPENDIX A.....	180
APPENDIX B.....	184
APPENDIX C.....	191
APPENDIX D.....	193
APPENDIX E.....	195
APPENDIX F.....	207

LIST OF TABLES

Table	Page
1. Comparing Results of the Pre- OPI with Post-OPI.....	111
2. Basem.....	113
3. Hazem.....	115
4. Ibrahim.....	117
5. Jamal.....	119
6. Salwa.....	120
7. Ramzy.....	122
8. Comparison of Language Features Means: Pre- and Post-Interventionist.....	124
9. Evaluative Feedback of the Pre-OPI and Pre-Interventionist.....	127
10. Evaluative Feedback between Post-OPI and Post-Interventionist.....	129
11. The Ten 5-Point Scales.....	136

LIST OF FIGURES

Figure	Page
1. Research Design.....	67

CHAPTER I

RESEARCH PROBLEM

Like a phoenix rising from the ashes, foreign language and intercultural education reached an unprecedented height of national importance as a result of the tragic and sad calamity of September 11th. Politicians, economists, and educators shockingly became aware of the US crippling lack of skills in foreign languages (Edwards, 2004). An astronomical number of positions in the government, the intelligence community, the private sector, and the military services unexpectedly erupted as a matter of national security (Kinger, 2002). Interculturalism became crucial to communicate with friends and enemies alike (Wesche, 2004).

This sobering awakening caused educators to search for the most effective ways of teaching foreign languages to adult learners (Brown, 2009; Kinger, 2002). They conducted several studies to identify an approach to second language acquisition (SLA) that would help students develop their skills rapidly to the highest proficiency level possible. Some of these studies found that integrating assessment into the process of language instruction to be effective or needed further investigation (Anton, 2009; Brown, 2009; Ellis, 2009a; Kinger, 2002; Lantolf & Poehner, 2009). Researchers used assessment to raise each learner's awareness to his or her needed language features, and this awareness helped students to internalize these new features promptly (van Lier, 1996). The results of the few studies available that investigated this approach on adult learners in one-on-one tutoring and classroom formats were promising (Ableeva & Lantolf, 2011; Allal & Pelgrims Ducrey, 2000; Anton, 2009; Brown, 2009; Dean, 2004; Poehner & Lantolf, 2005).

Statement of the Problem

Previous research studies showed dynamic assessment (DA) as a successful method for promoting second language acquisition and for measuring potential language learning (Anton, 2009; Dunn & Lantolf, 1998; Lantolf & Poehner, 2009; Poehner, 2005). Dynamic assessment was based on Vygotsky's (1978) concept of zone of proximal development (ZPD). ZPD is defined as the potential learning range in between each learner's independent and assisted performances (Vygotsky, 1978). Dynamic assessment identified the ZPD's borders for each learner by introducing gradual levels of assistance. These levels of assistance ranged from the most implicit to the most explicit standardized hints. Each level of explicitness was introduced only when the learner's existing abilities ceased to help the student produce a certain language feature as described on a particular scale (Poehner, 2005).

As mentioned above, these hints graduated from the most implicit hint such as not accepting the student's response to the most explicit hint by providing a full explanation for the answer. In between these two ends of a continuum, the teacher provided the student with three to four levels of assistance each of which became more specific about the uttered error until the teacher, mediator, provided the student with the correct answer along with its explanation. This type of mediation helped in diagnosing each learner's immature (incomplete) abilities by determining the level of explicitness for the assistance provided to perform a certain language feature (Poehner, 2005). These gradual levels of assistance promoted learning and details on their promotion of second language acquisition are provided later in the section on the theoretical rationale. Although these two dynamic-assessment capabilities of diagnosing learning needs and promoting second

language acquisition were considered empowering to the teaching and learning process, they needed further investigation to examine the practicality of dynamic assessment in a classroom setting of adult learners.

Most prior research studies (Hill & Sabet, 2009; Lantolf & Poehner, 2011; Poehner, 2005; Poehner & Lantolf, 2005) showed the positive effect of DA on instruction and demonstrated that the process was conducive to a prompt internalization of newly-learned language features. For example, if the student uttered the wrong word order in response to a question, the teacher would provide the first assistance by not accepting the answer. This hint might have led the student to reflect on his or her existing knowledge of the target-language grammar to produce the proper answer. The second level of assistance became more explicit than the first one by repeating the student's erroneous utterance, and then the teacher named the error by saying: "word order," for example. A more explicit level of assistance before providing the student with the answer could be by telling the student the proper word order in the target language without providing the answer. Later in the section on the theoretical rationale, the reader will find more details on how prompting the student to reflect promotes the internalization and improvement of immature abilities.

Most prior studies on dynamic assessment failed to address one or more of the following areas: (a) the use of dynamic assessment in language classrooms, (b) the instructional activities used, (c) the input materials used with students, and (d) the scale on which the dynamic-assessment process was calibrated. Previous studies (Lantolf, 2009; Poehner, 2005) that researched dynamic assessment in a tutoring setting with adult learners of French shed no light on activities used during these one-on-one tutoring

sessions. Other studies of adult students considered the cost-effectiveness of the classroom setting but did not document the activities used.

Adults are influenced by different principles of learning and teaching (Galbraith, 2004) as compared with children. Many dynamic-assessment researchers conducted their studies without fully documenting the teaching methods used with their adult participants, thus leaving the question of which methods would be most suitable for optimizing the use of dynamic assessment in adult classrooms unanswered. Other non-dynamic-assessment studies found that collaborative learning approaches such as task-based-language instruction, content-based instruction, project-based instruction (Stryker & Leaver, 1997), and performance-based assessment (Bachman, 1990; Galbraith, 2004b; M. H. Long, 2000; Messick, 1994) to be most suitable for adult learning (Brown, 2009; Galbraith, 2004a; H. B. Long, 2004). More precisely, the literature showed task-based language instruction to be one of the most effective approaches for second-language learning and teaching (Ableeva & Lantolf, 2011; Ellis, 2009a; Galbraith, 2004a; Nunan, 2004). One of the main principles of designing task-based language instruction was using suitable input material, and this factor was missing from the previous dynamic assessment studies as mentioned above.

Second language acquisition literature in general emphasized the importance of the relationship between the material used in language classrooms and the current proficiency level of the learners. Krashen (1987) emphasized the important relationship between the level of difficulty of the material presented to the learners in language classrooms and their current proficiency level. In his often-cited formula “ $i+1$,” he expressed the importance of using comprehensible input material that would be

understandable yet included one step higher to the learner's existing language abilities "i." Therefore, having a way of gauging the difficulty level of the oral and written text used in classrooms became fundamental to selecting the suitable material for the learners' current language abilities as well as their inabilities. Previous research not only failed to include the standards by which input material was selected but also did not document the scale used to evaluate their students' proficiency levels. In addition, it was imperative to use a particular scale for designing and calibrating rubrics that would describe the standards against which the learners' progression was measured and documented (Bachman, 1990, 2002; Bachman & Palmer, 1996; Brown & Abeywickrama, 2010). Previous research studies (Hill & Sabet, 2009) did not include such rubrics to evaluate the progress of students or to diagnose the language features needed for planning subsequent lessons necessary for students to progress on a particular scale.

Background and Need

The tragic and shocking attacks of September 2011 brought into focus the United States' deficit in foreign language capabilities. Politicians, economists, and educators realized the serious national need for efficient foreign language skills in the workforce (Edwards, 2004). People who could speak a language other than English were only 18% of Americans in 2010, whereas 53% of Europeans could converse in a second language (Skorton & Altschuler, 2012). The enrolment of Kindergarten to 12th grade (K-12) students in the year 2007-2008 reached 8.9 million (Skorton & Altschuler, 2012), and 1,682,627 students were enrolled in 2009 in courses for languages other than English in institutions for higher education (Furman, Goldberg, & Lustin, 2010). Although this number represented a slight increase over the number of students enrolled in 2006, the

increasing demand on learning foreign languages from 2002 to 2009 had become a recognizable trend (Furman et al., 2010; Skorton & Altschuler, 2012). This increase in foreign-language learning was noticeable for students who selected to learn Arabic. Arabic moved from the tenth place to the eighth spot for the most popular languages studied in the U.S. colleges (Furman et al., 2010).

The Secretary of Education, Arne Duncan, declared that foreign-language education was essential for the United States' economic growth and international relations (Skorton & Altschuler, 2012). The U.S. Government has been a key player in the movement of increasing and improving foreign-language education. The Foreign Service Institute (FSI), which is the Department of State's source for foreign language education, train about 100,000 individuals in more than 70 languages a year (FSI, 2013). The Defense Language Institute Foreign Language Center (DLIFLC) educates about 3,500 students in more than two dozen languages each year (DLIFLC, 2013c). This increasing national demand for foreign language learning and in the U.S. Government in particular has been growing in shrinking budgets since the infamous financial crisis of 2007. Studies had been conducted in the Federal Government and elsewhere to improve foreign language instruction (Brown, 2009; Gnadinger, 2008) in the classrooms. Dynamic assessment emerged from these studies as a plausible and a cost-effective approach for improving foreign-language instruction.

All studies found on dynamic assessment (Anton, 2009; Dean, 2004; Lantolf & Poehner, 2009, 2011) did not include the principles of adult learning. Most of them were conducted in one-on-one tutoring sessions (Poehner, 2005). These studies neither used a particular proficiency scale to evaluate the students' progress nor used particular rubrics

for the daily dynamic-assessment activities in classrooms or in the one-on-one format. Very few studies combined collaborative learning techniques such as task-based language instruction, project-based instruction, or content-based instruction with dynamic assessment for gauging an individual's or group's ZPD while solving a problem in a real-life situation (Anton, 2009; Brown, 2009; Doolittle, 1995, 1997). All previous studies failed to investigate the dynamic-assessment approach for teaching to adult learners in a classroom setting or in tutoring sessions. Arabic is very important for the US Government in general and consequently for the Defense Language Institute Foreign Language Center (DLIFLC) in particular.

DLIFLC had not conducted any formal study on using dynamic assessment in their Arabic program, even though dynamic assessment had been on the rise since the 1970s (Carlson & Wiedel, 1978). Although dynamic-assessment researchers had incorporated Vygotsky's (1978) ZPD in many studies, traditionalists were still resistant to the idea (Kinging, 2002). This tension in the educational vision invited an investigation by Kinginger (2002) to find out if mediation in the ZPD was conducive to second language acquisition. The U.S. foreign language profession needed to determine how to exploit the "dialectical process" (Kinging, 2002, p. 257) in the ZPD in a broader sense. Kinginger (2002) used the term dialectical process to refer to the dynamic-assessment approach, because it combined assessment and teaching in the language teaching process. Dynamic assessment had been known as a monistic approach (Poehner, 2005), because it combined assessment and teaching in the same activity like two sides to the same coin.

Kinging (2002) found that a broader understanding of using the ZPD in collaborative activities would help in advancing the agenda of communicative language

teaching. Collaborative learning could be accomplished easily when designing a task-based-language-instruction activity by prompting learner's to cooperate experientially in small groups. This process needed practical operationalization that would solve the long tension between progressive language education and conservative educators. Kinginger (2002) expressed the importance of defining the term "effectiveness" so that the result of co-authoring and co-constructing in an experiential real-life activity would be measurable. This current study contributed to the definition of "effectiveness" by using the Interagency Language Roundtable scale to measure the progress of students during the DA activities.

This current study combined task-based language instruction with dynamic assessment in a classroom setting of adult learners of Arabic, using validated rubrics that were rooted in the U.S. Government's scale for evaluating foreign language proficiency, the Interagency Language Roundtable scale. The difficulty level of the input material used by the researcher in the classroom was measured by the principles of text typology (Child, 1987, 1998, 2001) as used at the Defense Language Institute Foreign Language Center (DLIFLC, 2013c, 2013d). Therefore, this study investigated the effectiveness of combining dynamic assessment and task-based language instruction for Arabic speaking as a second language, and it explored the nature of teacher experience and perception relative to the implementation of dynamic assessment and task-based language instruction for Arabic as a second language.

Purpose of Study

The purpose of this study was to investigate the effectiveness of combining dynamic assessment with task-based activities that targeted the speaking skill of Arabic

(Goos, Galbraith, & Renshaw, 2002; H. B. Long, 2004) and task-based language instruction that included small-group collaborations in Arabic for the purpose of creating measurable products. More specifically, this dissertation explored the effect of using an ongoing classroom assessment (Anton, 2009; Bachman, 1990) to gauge and exploit Vygotsky's (1978) ZPD of each learner or a group of adult students of Arabic (Allal & Pelgrims Ducrey, 2000; Dean, 2004). Providing instruction through gauging and scaffolding into learners' ZPD was known in the field of foreign language education as dynamic assessment (DA). This mixed-method study was designed to contribute to the knowledgebase developed from previous studies of the effectiveness of DA-based instruction.

It investigated the practicality of continually assessing students' weaknesses and strengths during their course of instruction and particularly as a group (Brown, 2009; Ellis, 2009a). This research was designed to use the proficiency scale used in the U.S. Government with students attending the Defense Language Institute Foreign Language Center. These students were military service men and women, and they were attending the Arabic Basic Course. Therefore, the findings of how effective dynamic assessment was in their daily classroom instruction could benefit language-adult-learning programs at DLIFLC, colleges, universities in the United States and around the world.

This dissertation investigated the effect of combining task-based language instruction in classrooms with dynamic assessment on the students' Arabic speaking abilities. The process of combining both these approaches was referred to in this study as DA/TBLI instruction. The process of DA/TBLI instruction was guided and measured by the U.S. Government's proficiency scale known as the Interagency Language Roundtable

scale. The study used Interagency-Language-Roundtable-based rubrics guided by a table format found in performance-based assessment (Johnson, Penny, & Gordon, 2009). The standards for the different targeted independent performances for students were established for this study by deconstructing the Interagency-Language-Roundtable scale into recognizable sublevels for the ranges between the description of every two existing proficiency levels (ILR, 2013a).

These recognizable sublevels helped fulfilling the study's purpose, because they provided a valid and reliable measuring instrument for gauging the effect of dynamic-assessment-based instruction on both language learning and students' diagnosing. The study's rubrics measured the effect of dynamic assessment on language learning and on the diagnosing ability for students' needs. The Defense Language Institute had been using task-based language instruction in its language-teaching programs for over 10 years and a process called Diagnostic Assessment for about 15 years. The Arabic schools used mainly Diagnostic Assessment with some students two times during its Arabic Basic Course. The daily process for diagnosing students' needs had been accomplished mostly by the teachers' personal observation or by conducting Oral Proficiency Interviews during the programs formative-assessment tests. The schools provided students with periodic Oral proficiency Interviews toward the end of the basic course prior to their formal exit test.

Research Questions

Studying the effect of DA/TBLI was measured by using the Interagency-Language-Roundtable rubrics to investigate the change in the students' performance at the end of this study. To make this measuring more practical for the purpose of this study,

the focus was only on one accuracy factor for each proficiency level of the Interagency Language Roundtable scale. The accuracy factor measured in this study was the “structural control.” To measure the effectiveness of the DA/TBLI approach on adult learners of Arabic, this study addressed the following research questions:

1. What is the change in the structural control of Arabic speaking based on DA/TBLI instruction?
2. How do OPI without DA assistance and OPI with DA assistance compare relative to the evaluation of Arabic speaking?
3. How do the experiences and perceptions of DA/TBLI instruction compare between teacher-researcher and OPI testers?
4. What are the student perceptions of the DA process?

Theoretical Rationale

This proposed study was based on two theoretical models: sociocultural theory and task-based language instruction as a suitable approach for adult learners. The first theoretical model was Vygotsky’s (1978) sociocultural theory. According to the sociocultural theory, development occurs through social co-construction of meaning within an area that stretches between the child’s assisted and independent performances (Vygotsky, 1978). This area, the ZPD, has been used to identify the learners’ needs through observing the type and level of assistance necessary for learners to perform a given language task (Anton, 2009; Poehner, 2005) as described on a particular scale.

The collaboration and the guidance mentioned above were used to measure the learner’s area of ZPD and to identify its borders. This mediation was the key part of DA; “We first define the theoretical concept of DA based on Vygotsky’s ZPD, which

integrates mediation and assessment into a unified pedagogical activity” (Ableeva & Lantolf, 2011, p. 133), that is the teacher becomes a mediator between the student’s current ability and the desired performance of the targeted language feature or task as required on a particular scale (Poehner, 2005). In this study, this mediation helped the teacher-researcher to realize the distance between the learner’s current immature abilities and the needed independent performance as described on the Interagency Language Roundtable (ILR) scale for the targeted proficiency level.

The assistance provided during the mediation not only could identify the learner’s immature language features but also would promote their development (Poehner, 2005). Providing the missing information to students while being focused completely on their language weakness as needed to convey their thought would maximize their awareness of the discrepancy in their language abilities. Therefore, this heightened state of awareness would lead not only to their fast internalization of the linguistic element at hand but also to their independent performance of it quickly (Poehner, 2005; van Lier, 1996). Consequently, it was assumed that the language feature that a student could perform initially only with assistance would soon be performed independently to meet the targeted descriptors of the used scale (Poehner, 2005). This progress as measured by the used scale would help the learner to discover other language features, and the reiterative cycle of dynamic assessment would promote the learner’s progression on the used scale effectively (Poehner, 2005). Using dynamic assessment in this study reflected my advocacy for the sociocultural theory in classrooms.

The second theoretical model was the suitability of task-based language instruction for adult language classrooms (Ellis, 2009a, 2009b; Foster & Skehan, 1999;

M. H. Long, 2000; Skehan, 1998; Skehan & Foster, 1999). Task-based language instruction (TBLI) activities are student-centered (Bachman & Palmer, 1996) and effective for adult-learning (Bachman & Palmer, 1996; Dean, 2004). As prescribed for adult learners (Galbraith, 2004b), these activities prompt students to think critically for the purpose of solving a real-life situation by generating a measurable product. This generation of deliverable products led students to use their background knowledge and differences in the productive modes of the target language. While students would be working on solving the problem, two different ZPDs would emerge.

The first ZPD was the area between peers during their group work, and the second area started between the group's collective language ability and the teacher-researcher's when they presented their group product. This second situation was known as the Group-ZPD, and it allowed the researcher who was also the teacher in this study, henceforth referred to as the teacher-researcher, to use the same concept of dynamic assessment with the whole group (Brown, 2009; Hill & Sabet, 2009). Task-based language instruction was used in all the classroom activities of this study, and their efficacy when combined with dynamic assessment was measured by rubrics devised for this study. These rubrics were based on the speaking section of the ILR scale (ILR, 2013a) for evaluating and recording the students' progress. The suitability of combining dynamic assessment with task-based language instruction for adult learning was very important for this study. Task-based instruction allowed adults to incorporate their knowledge of the world and different personal profiles to think critically for the practical purpose of solving a real-life situation, and these hands-on and progressive way of learning were highly prescribed for adult learners (Dean, 2004; Dewey, 1963). To maximize the effect of the task-based

activities used in this study, the teacher-researcher was mindful of the different variables among adult learners.

Adult learners and teachers vary in many ways that effect the learning and teaching process dramatically (Galbraith, 2004b). Their variability would be caused by life experiences, which would make a clear distinction between each person's brain and mind (Bialystok, 1994). The teacher-researcher believed that the mind would start in the memory stored in the brain and it would extend to include all the surroundings of each person (Piaget, 1971), which would make every adult unique in learning and processing information. Each person's mind would be limitless in size to include all the knowledge gained from books, trips, schools, and people met, and the mind stays expandable to new dimensions into the future. This future would depend on the intervention that takes place through learning experiences and social interactions (Poehner, 2005).

A later section on dynamic assessment will handle these new dimensions into the future of learning, but this section continues to address the diversity of adults' minds. The teacher-researcher also believed that a person's mind would store different information, memories, understandings, epistemological convictions, and world views because of the different social stimulants and triggers that they had experienced in their lives growing up. Adults would be more diverse physiologically, psychologically, and sociologically than children (H. B. Long, 2004). Their psychological differences would include cognitive, personality, and experiential and role characteristics.

The cognitive characteristics would reflect the level of maturity for a person, and it has four stages (Piaget, 1971). According to Piaget, these four stages were: sensory-motor stage (to about 2 year), pre-operational stage (2 to 6 years), operational stage (7 to

11 years), and formal stage (12-15 years). The formal-operations stage was referred to as the abstract level. Piaget (1972) raised that limit to 20 years of age and researchers found later that age alone would not guarantee attainment of formal stage operational abilities (H. B. Long, 2004). The age inability of helping a person reaching the formal stage simply would mean that adult learners would not be at the same cognitive level, which would affect directly their reaction to learning experiences. The other factor was their personality characteristics.

Long (2004) reviewed the literature and reported that personality was defined as the consistent way of behaving, and it had eight multidimensional properties: (a) physique, (b) temperament, (c) intellectual and other abilities, (d) interests and values, (e) social attitudes, (f) motivational dispositions, (g) expressive and stylistic traits, and (h) pathological trends. Each one of these factors would affect adults in the way they approach learning at large or a specific learning situation (H. B. Long, 2004). The reason would be that each one of these factors would include a wide variety of levels and types, and, therefore, treating all adults as if they follow a specific prototype would be a faulty assumption. Teachers of adults ought to consider the above-mentioned variables in their lesson planning, classroom interactions, and dividing students into small groups, because each learner would come to class with her or his unique idiosyncrasies, cognitive, and learning style.

Learning styles would be the results of the unique profile of each adult learner and their personal history, child development, cognitive development and all the variables mentioned in the last paragraph. The term “style” caused a controversy and disagreement in the field of education for its overlapping proximity with personality and ability (Zhang,

Sternberg, & Rayner, 2012). They recognized three reasons for the challenges that were facing learning styles. “We see the field as having been presented with three principal challenges: (a) a lack of identity, (b) the existence of three major controversies concerning the nature of styles, and (c) the confusion brought about by several critical reviews of the field” (Zhang et al., 2012, p. 2).

They explained the lack of identity as the direct consequence of using overlapping terms that were synonymous to the term “learning styles” such as “cognitive styles,” “thinking styles,” “mode of thinking,” “mind styles,” and “teaching styles” (Zhang et al., 2012, p. 1). To avoid this problem, Zhang & Sternberg(2005, p. 1) coined the term “intellectual styles” as an overarching term that covered all the source philosophies of the other terms. This present study identified the different intellectual styles and different background information of participants in order to maximize interactions among its participants during the task-based activities.

Definition of Terms

This section provides the meaning of some terms as intended and used in this study.

Defense Language Institute Foreign Language Center (DLIFLC): the Defense Language

Institute Foreign Language Center is the foreign language teaching school in the United State’s Armed Forces (DLIFLC, 2013c).

Dynamic Assessment (DA): This type of assessment is designed to gauge learners’ potential development by providing students with various levels of scaffolding and it is based on Vygotsky’s zone of proximal development (Grigorenko & Sternberg, 2002; Poehner, 2005). This study used dynamic assessment to

diagnose students' needs in its initial phase and to promote language acquisition in the classroom.

Dynamic Assessment Rubrics Form (DARF): DARF is the form devised for this study for the purpose of operationalizing dynamic assessment during classroom activities. It was developed by deconstructing the Interagency-Language-Roundtable scale (ILR, 2013a) for the range of proficiency levels expected for the participants of this study.

Interactionist DA: This is the type of dynamic assessment that is used usually during language instruction for the purpose of promoting second language acquisition (Grigorenko & Sternberg, 2002). This DA type was used to promote language acquisition in the classroom activities.

Interagency Language Roundtable (ILR): This is a proficiency-based scale used in the U.S. Government to evaluate foreign language abilities in Listening, Reading, Speaking, and Writing (DLIFLC, 2013b).

Interventionist DA: This is a dynamic-assessment interview used in a pretest-posttest format for foreign language diagnostic and learning purposes (Grigorenko & Sternberg, 2002). This DA type was used in this study to diagnose the students' needs during its initial phase.

Oral Proficiency Interview (OPI): The OPI is a psychometric (static) speaking test in a foreign language and it is used as a summative evaluation at the Defense Language Institute Foreign Language Center (DLIFLC, 2013b). OPIs were used at the beginning and end of this study to measure improvement in proficiency.

Task-Based Language Instruction (TBLI): TBLI is an approach for students to work collaboratively using a multimodal input to generate an observable product that solves a real-life situation (Ellis, 2009b). TBLI activities were combined in this study with DA interactionist (DA/TBLI instruction) in the classroom.

Zone for Proximal Development (ZPD): The ZPD is the learner's mental area bordered between his or her assisted and independent performances (Vygotsky, 1978). The students' ZPDs were used in this study through a scaffolding process of gradual hints.

Significance of Study

Finding the answers to the study's questions would not only contribute to the language learning process at DLIFLC but also may eventually contribute to improving adult foreign language learning and teaching in the US and possibly worldwide. Having students daily for 3 to 4 weeks toward the end of DLIFLC's unique 63-week Arabic Basic Course who volunteered for this study could inform the practice of foreign language learning and teaching of adults. Having professional military students who were motivated to further their objectives eliminated many negative learners' variables that could have existed in other adult programs. Their constantly updated curriculum supplemented dynamically to stay abreast with the latest in the field provided this study with students who were used to task-based activities and to being immersed in the realistic uses of Arabic.

Classrooms were empowered by state-of-the-art technological and networking resources; the advanced technological resources available made simulating real-life situations in task-based activities much easier than other programs. Unlike most

programs, students were in this immersive environment daily for 7 hours driven by the proficiency-based-ILR descriptors. Conducting this study in real-life-simulated activities assisted the combining efforts of dynamic assessment with task-based language instruction effectively. Consequently, this study may encourage more studies in the future for the purpose of generalizing its findings. This generalization might contribute to the practice of foreign language classrooms in colleges and universities around the globe.

The next chapter reviews the literature for dynamic assessment and task-based language instruction. Each topic section includes background information and a review of previous studies.

Chapter III presents the methodology used to answer the questions of this study. It includes the research design and data analysis techniques used. Chapter IV provides findings that answer the study's questions. Finally, Chapter V offers a discussion of findings and a conclusion. This conclusion leads to recommendations for practice and future research suggestions.

CHAPTER II

LITERATURE REVIEW

The purpose of this study was to investigate the effectiveness of combining dynamic assessment with task-based activities that would target the speaking skill of Arabic (Goos, Galbraith, & Renshaw, 2002; H. B. Long, 2004); Task-based language instruction (TBLI) activities included small-group collaborations in Arabic for the purpose of creating measurable products. More specifically, this dissertation explored the effect of using an ongoing classroom assessment (Anton, 2009; Bachman, 1990) to gauge and exploit Vygotsky's zone for proximal development (ZPD) of each learner or a group of adult students of Arabic (Allal & Pelgrims Ducrey, 2000; Dean, 2004). Providing instruction through gauging and scaffolding into the learners' ZPD is known in the field of foreign language education as dynamic assessment (DA). This mixed-method study was designed to contribute to the knowledgebase developed from previous studies of the effectiveness of DA-based instruction, and this chapter reviews studies on dynamic assessment and task-based language instruction. The section on dynamic assessment will include important concepts of DA, the ZPD and its use in DA, both types of DA, and previous DA studies. The section on TBLI will include background on performance-based assessment and previous TBLI studies.

Dynamic Assessment

Dynamic assessment as an evaluation instrument and a learning approach was based on a compelling logic and a sound theory for education at large and recently for foreign language learning in particular. As an assessment instrument, it was used for diagnosing the strengths and weaknesses of a foreign or a second language learner at the

beginning and the end of language-training programs (Poehner, 2005). It was used also as a learning approach of second language acquisition and specifically in one-on-one tutoring sessions (Poehner, 2005). Previous studies on dynamic assessment focused mainly on exploring the effect of this approach on second language learning in a tutoring context. Dynamic assessment had a variety of measurement techniques known by certain labels. Examples of these techniques were testing the limit (Carlson & Wiedel, 1978), learning potential assessment (Budoff, 1987a, 1987b), and learning tests (Guthke & Stein, 1996). These techniques shared a common feature of having an element of teaching in the form of examiner intervention. Tutoring, coaching, or mediation were integrated in the assessment sequences for the purpose of obtaining better evaluation of the learner's cognitive abilities and more accurate prediction of his or her potential learning (Allal & Pelgrims Ducrey, 2000).

To make this concept marketable and cost effective, it had to be operationalizable effectively in a classroom setting of second language learning. This literature review begins with the literature on the preexisting knowledge in the field about using dynamic assessment in a classroom environment and not only for tutoring purposes. To be more specific, the purpose of this literature review is to explore DA's effectiveness when used with a group of adult learners in a classroom setting. Dynamic assessment was based on Vygotsky's (1978) sociocultural theory, and it is defined usually in a tutoring context. Therefore, the first question of this study was about the change that would occur as a result of combining dynamic assessment with task-based language instruction in classroom activities.

To answer this question, the first section begins by identifying the common definitions of terms and the operational process of dynamic assessment. To that end, this literature review of dynamic assessment presents first the studies and articles for the relevant information available for those terms and definitions. Then, the second and the following section will review the previous studies of dynamic assessment for both evaluation and learning purposes. These studies included dynamic assessment, peer-assessment, and collaboration as possible components of using dynamic assessment in a classroom setting. The information found serves as foundation for the activities used in this study as explained in the next chapter.

Important Concepts of Dynamic Assessment

Traditional testing or known henceforth as static assessment separated testing from learning completely. The main purpose of static assessment tests was measuring present abilities at a certain point in time (Brown & Abeywickrama, 2010; Poehner, 2005). If the evaluation of abilities were done at the end of a certain course of instruction, curriculum, or program, then the test would be known to be a summative assessment, and it would be called a formative assessment when administered during the course of instruction (Bachman, 2002; Brown & Abeywickrama, 2010). Formative assessment would be designed usually to determine whether a learner was on track toward the end objective of a language program. This evaluation of a student during the course of instruction would reflect the learners' abilities of mastering the material covered during the preceding period in the program. If the results of a formative test would affect subsequent classroom instruction, then the formative test would be high on "consequential validity" (Brown & Abeywickrama, 2010). Unlike static assessment,

dynamic assessment could guide effectively subsequent lesson planning due to its diagnostic ability for immature abilities during the daily course of instruction (Poehner, 2005).

The issue with static tests as a method of formative evaluation was that they only measure the existing mature abilities, but they were unable to identify any knowledge or skill that was still in the making (Poehner, 2005). Static tests were unable to inform foreign language educators about how far away a learner was from performing a language feature independently. Vygotsky's ZPD exploited the learner's needs for assistance to diagnose the abilities that were still in the making (Grigorenko & Sternberg, 2002). This concept was found by researchers (Ableeva & Lantolf, 2011; Anton, 2009; Budoff, 1987a, 1987b; Carlson & Wiedel, 1978; Guthke & Stein, 1996; Lantolf & Poehner, 2011) to be effective in measuring both mature and immature abilities through a mediation process conducted by the teacher (Poehner, 2005). Teachers who used dynamic assessment played a dual role of being instructors and testers at the same time. When they provided their assistance for the purpose of diagnosing and evaluating the students' abilities and inabilities, they were called mediators. The mediation process in the learner's ZPD was known as dynamic assessment. Before exploring the effectiveness of mediation on the learning process, the next section elaborates further on ZPD. ZPD was at the heart of dynamic assessment, and measuring it and working in it was imperative in answering the first question of this research.

ZPD and Its Use in Dynamic Assessment

First, Vygotsky (1978) explained that the ZPD is the area between a learner's assisted and independent performances, and he stated that the ZPD would be the distance

between the actual development level as determined by independent problem solving and the level of potential development as determined under adult guidance or in collaboration with more capable peers. Based on this type of collaboration, the mediation mentioned above was used to measure the learner's area of ZPD. This mediation was the key part of dynamic assessment (DA). The theoretical concept of dynamic assessment was based on Vygotsky's ZPD, which integrated mediation and assessment into a unified pedagogical activity (Ableeva & Lantolf, 2011). Dynamic assessment combined both learning and testing in the same instructive activity by assisting the student while attempting to perform the language needed to fulfill a certain task (Allal & Pelgrims Ducrey, 2000). The teacher or the tester became a mediator between the student's current ability and the desirable performance of the targeted language feature.

This mediation was the key of the DA process in the learner's ZPD as explained in this citation: "DA requires the examiner to mediate the examinee's performance during the assessment itself through the use of prompts, hints, and questions" (Poehner, 2005, p. iii). Some researchers used gradual and standardized hints to measure immature abilities and how far the learner was from performing independently (Poehner, 2005). Gradual standardized hints are found in Poehner (2005, 2010) as graduating from the most implicit to the most explicit. In the procedures developed by others (Budoff, 1987a; Ferrara, Brown, & Campione, 1986; Guthke & Stein, 1996; Lidz, 1991), a standardized sequence of general to specific prompts or hints was proposed, and these standardized number of hints graduated from the most implicit to the most explicit hints (Poehner, 2005). A student who was close to perform a certain language feature independently needed a few implicit hints while a weaker student needed more explicit hints.

Each level of this graduation was provided only when the learner's own abilities ceased to be of help to her or him. By identifying the level of explicitness, the mediator could measure the level of maturity for abilities that were still in the making. The mediator could identify precisely which language features and information were needed for the learner to reach the desired independent performance. The DA process enabled the mediators to evaluate the person's abilities and immature abilities, which was more information obtained than what the traditional static assessment could measure or provide (Grigorenko & Sternberg, 2002). Due to the entrenched Western traditions and convictions in regard to the validity and the reliability of a test (Bachman, 1990, 2002; Brown & Abeywickrama, 2010), dynamic assessment could only provide a diagnosis of a learner's existing abilities and potential learning abilities. That is dynamic assessment would not be a replacement for traditional testing. "The findings suggested that DA would be an effective means of understanding learners' abilities and helping them to overcome linguistic problems" (Poehner, 2005, p. iv). Poehner (2005) conducted his study in a tutoring setting for six students in one-on-one sessions. The first two questions of this dissertation investigated rather the effectiveness of Poehner's findings in a classroom setting for Arabic. The next two sections present the two types of DA used in this study, DA interventionist and interactionist.

DA's Interventionist Model

This diagnostic approach is known as the "interventionist approach" of dynamic assessment (Poehner, 2005, p. 22), and usually this approach was used in a pretest-posttest format. Learners received their first dynamic assessment process before the beginning of a language program, and based on the findings of this interview, a tailored

program was designed for the learner. This program was called the treatment in most previous studies, and dynamic assessment was used also in the daily instruction of this study. In this study, the term instruction means the treatment program of previous studies. The results of the posttest conducted at the end were compared with the results of the first test to identify the student's accomplished progress. A more accurate descriptive naming of this DA approach was test-teach-test. Both tests were designed usually to measure the same features and have similar structures. The current study combined the structure of the Oral Proficiency Interview with the techniques of the interventionist DA. Not only dynamic assessment was used in the pretest-posttest approach for diagnosing and measuring the learner's needs and progression at the beginning and the end of a language program, dynamic assessment was used as well for the daily instruction in between these two interventionist-DA interviews. The dynamic assessment used during the daily instruction in between the pretest and posttest interviews was called interactionist DA (Grigorenko & Sternberg, 2002; Poehner, 2005). Interactionist DA followed the same concept of foreign language instruction that was based on Vygotsky's ZPD (Poehner & Lantolf, 2010).

DA's Interactionist Model

The approach, known as dynamic assessment (DA), a term coined by Luria (1961), derived from Vygotsky's own work in the area of "defectology" and aimed at reveal abilities that fully developed as well as those that were still forming (Poehner, 2005). The other side of this "dualistic" approach was learning. Learning occurs when a person would interact with a stronger peer or a teacher who would assist the learner in overcoming a certain difficulty. This development took place only when learners could

not depend on their existing abilities or knowledge to perform independently. Based on Vygotsky's (1978) sociocultural theory, social interaction provided learners with the needed trigger mechanism to activate their own cognitive process (Lantolf & Poehner, 2009).

Social interaction would allow the learner to connect the newly received knowledge with their existing abilities to progress into a more complex and advanced performance, knowledge, or understanding. "DA techniques provide learners with a 'mediated learning experience' (Lids 1991, p. 14) in which, through social interaction, experiences are filtered, focused, and interpreted as needed by the learner" (Anton, 2009, p. 579). The social engagement with the learner's cognitive process would allow the gaining of the new information by making sense of the unknown part in terms of their existing knowledge. Vygotsky (1978) expressed his theory by stating that today's assisted performance would be tomorrow's independent ability and the difference between the two would be the learner's potential learning (Poehner, 2005).

This side of the dynamic-assessment process was termed the "interactionist model" (Poehner, 2005, p. 161). Dynamic Assessment was a dialectical approach in reference to learning a second language, because it used both assessment and instruction for foreign language acquisition. For the same reason, it was also called a "monistic approach" because both assessment and instruction were used inseparably like two sides of the same coin for foreign language learning (Poehner, 2005, p. 151). Logically and for the purpose of this literature review, the interactionist technique was the focus for the remaining parts of this review. "Vygotsky's theory, variously referred to as cultural historical or sociocultural theory, proposed that human development would arise from the

dialectical interaction of lengthy biological evolution and sociocultural changes propagated over the course of human history” (Ableeva & Lantolf, 2011, p. 133). This statement inspired the teacher-researcher to choose task-based language instruction for creating social venues combined with dynamic assessment in a classroom setting.

Transfer of Learning

Although Vygotsky’s sociocultural theory was a psychological theory and was not intended for second or foreign language learning and teaching, the importance of social interaction was common in both his work and the mainstream of the second-language-acquisition field. Although Lantolf (2012) rejected Chomsky’s (1968) famous Language Acquisition Device and expressed his disbelief in its existence in his presentation at the annual convention of Teachers of English to Speakers of other Languages, both the sociocultural theory advocates and the mainstream SLA theorists including Lantolf believed in the importance of social interaction for cognitive development. Regardless of any possible theoretical conflict (Poehner & Lantolf, 2010), the human innate ability to learn a language through cultural interaction was still a commonly held belief in the work of Vygotsky, Piaget, Luria, Poehner, Lidz, Budoff, Guthke and mainstream SLA writers (Bialystok, 1994; Larsen-Freeman, 1991b; Swain, Kinnear, & Steinman, 2010; van Lier, 1996).

As for the Language Acquisition Device, the evidences for the human innate ability to speak a language was overwhelming (Chomsky, 1968), and social interaction solely was insufficient to learn a language. The impact of social interaction as the trigger mechanism for the activation of this biological built-in ability had been investigated by many scholars (Bialystok, 1994; Canale & Swain, 1980; Larsen-Freeman, 1991b; Swain

et al., 2010). Covering the importance of social interaction would need its own paper, and the scope of this study was mainly DA. Therefore, considering that dynamic assessment was based on Vygotsky's (1978) sociocultural theory, the importance of social interaction and the different components of communicative competence (Canale & Swain, 1980; Swain et al., 2010) were in compliance with the mainstream of the field of second language acquisition. Not only that the DA-provided social interaction would be conducive to the learning of a new language feature, but also it would be crucial in developing the learner's proficiency in the target language.

Proficiency was measured in the Interagency Language Roundtable scale by descriptors sorted in the following categories: (a) lexical control, (b) grammatical control, (c) sociocultural competence, (d) delivery, (e) text type (length of utterances), and (f) global tasks. These categories were congruent with the factors mentioned for communicative competence (Canale & Swain, 1980). Canale and Swain's (1980) factors were (a) grammatical competence, (b) sociolinguistic competence, and (c) strategic competence. Sociolinguistic competence in the Canale-Swain (1980) model of communicative competence could be broken down into two kinds of competence: sociocultural competence and discourse competence.

Going through a graduation of complexities reflecting the different components of communicative competence could be done by the transferring of a newly learned language feature to different situations and contexts. For example, the learners would develop their ability to perform a certain language feature independently through the provided gradual DA assistances. Then, the transfer of learning process would help them to use the same language feature appropriately in different cultural contexts. These

cultural contexts would need to graduate in complexity toward the targeted descriptors of the objected proficiency level of a certain scale or guidelines such as the Interagency Language Roundtable (ILR, 2013a) and the American Council on the Teaching of Foreign Languages (ACTFL, 2012) scales.

Both the interactionist and the interventionist models would measure and use the transfer of the targeted language features to different contexts as part of the complexity graduation needed until the learner's performance meets the standards of the assessed descriptor. Integrating this technique into the DA process in its interactionist model would cause the learner a deeper processing of the language features in question (Anton, 2009). The transfer-of-learning process would be a meaningful strategy for the development of a certain language feature toward the learner's performance of it independently (Hill & Sabet, 2009). Reproducing a certain language feature independently, properly, and suitably in all applicable situations would be logically a much higher ability than being able to reproducing the same language feature only in a simple context.

Previous DA Studies

In his lengthy dissertation on the effect of dynamic assessment on oral proficiency among advanced second-language learners of French, Poehner (2005) conducted a study on six participants. These participants were students in the advanced French program at the Pennsylvania State University. Poehner's (2005) extensive literature review explored all the previously used techniques of dynamic assessment. The study's questions focused on (a) the possibility of dynamic procedure adding to the understanding of the individual's knowledge of and ability in the second language, (b) the extent to which

interactions during dynamic assessment would promote learners' development, (c) the effectiveness of insights into learners' abilities gained from DA in developing an enrichment program that would tailor instruction to the individual's abilities and weaknesses, and (d) the possibilities of changes that would occur in the participants' performance during the course of enrichment (instruction) while performing tasks beyond those used for the initial assessments. To answer these questions, the researcher followed a test-enrichment (instruction)-retest approach.

At the beginning and at the end, each participant went through a static test and a dynamic assessment that were called Time 1 and Time 2, and the instruction of dynamic assessment was introduced in one-on-one tutoring sessions in between Time 1 and Time 2. The initial Time 1 tests were referred to as Static Assessment 1 and Dynamic Assessment 1, and the posttests (Time 2) were referred to as Static Assessment 2 and Dynamic Assessment 2. For Static Assessment 1 and Dynamic Assessment 1, students watched a video clip and then narrated the scene in French. The results of these assessments were used then to structure the instruction program; these diagnostic feedbacks from the Dynamic Assessment 1 provided insights into (a) the kinds of problems learners encountered while completing the tasks and (b) the amount and quality of collaboration with the mediator they required in order to overcome these problems.

Then after the instruction program, students went through Time 2 (Static Assessment 2 and Dynamic Assessment 2) during which the initial assessment was repeated. In Time 2, students received a "transfer assessment" (Transfer 1 and Transfer 2), and both were conducted to understand the extent to which participants could extend their learning beyond the original assessment context. Students went through all the

following developmental and mediated assessment programs: Dynamic Assessment 1, Dynamic Assessment 2, Transfer 1, Transfer 2, their own instructional school course, and the instruction of one-on-one tutoring program offered by the study. Six of those students volunteered at the beginning, but then only four of them participated in Time 1, Time 2, and the instruction program. Another two students participated only in Time 1 and Time 2.

Using Vygotsky's (1978) definition of development as "conscious awareness," Poehner (2005) justified its occurrence and nonoccurrence with both instruction students and noninstruction students (students who were not interacting with the teacher or the stronger peer). The instruction students are the ones interacting with the teacher or the stronger peer in or outside of classrooms. Through analyzing the data in three chapters, Poehner (2005) found that development occurred due to both kinds of mediation: the "cake/interactional," and the "sandwich/interventionist." The following is a more detailed answer for each research question. As for question number one, static assessment was found as expected to be capable only of measuring independent performance, but also only dynamic assessment was able to measure immature abilities (abilities that are still in the making).

The second question was answered through the participants' verbalization. Poehner (2005) used a participant named Nancy to show how assistance during assessment would cause development. Poehner (2005) expressed that it was safe to conclude that the change in Nancy's performance at Time 2 was, in large measure, the result of her interactions with the mediator during Dynamic Assessment 1. The third question was about individualizing instruction. Poehner (2005) repeated his explanation

about the Learning Potential Measure and mentioned that dynamic-assessment researchers, such as Feuerstein, argued that static procedures do not reveal the underlying sources of poor performance and only reinforce learner's frustration with assessment. Then, he mentioned that several insights into learners' abilities were gained only through interaction during Dynamic Assessment 1 and Dynamic Assessment 2.

Poehner (2005) did not mention the scale on which he evaluated the learner's progress. In these tutoring sessions, it was obvious that the researcher provided the gradual standardized hints to assist the learner overcoming the initially-diagnosed grammatical features. The reader of this research was left to wonder about the importance of these grammatical features. Were they important for passing certain standards for the final examination of the students' advanced French class? What were the criteria for being advanced in French? By which scale students were measured as being advanced? Were students evaluated by an achievement-based scale or by a proficiency-based scale (Defense Language Institute Foreign Language Center, 2010; Interagency Language Roundtable, 2012a)? Poehner's (2005) study did not include the activities used with its participants for the readers to know whether they had any contributions to the results or whether they were usable in a classroom setting.

Poehner (2005) conducted his research in a tutoring format, which left the readers of his study questioning the practicality of using dynamic assessment in a classroom setting. This current dissertation replicated several designing aspects of Poehner's (2005) study, but it included additionally the combining of task-based language instruction and dynamic assessment as the teaching approach used in its classroom sessions. In regard to the lack of using a particular scale in Poehner's (2005) study, this current dissertation

avoided this shortcoming by using the Interagency Language Roundtable scale. The operationalization of all these factors in a classroom setting prompted the questions of this current research.

Unlike Poehner's (2005) study, Gnadinger (2008) conducted a study using DA in a classroom setting. The focus of his study was on peer-mediated instruction and the assisted performance in a classroom. Gnadinger's (2008) study was conducted on multi-age primary classroom in the Southeastern region of the United States. Gnadinger (2008) studied the ways elementary-school students provided scaffoldings to one another while immersed in collaborative activities. These students were second and third graders who ranged in age from 7 to 9 years of age. This study supported findings that students while interacting to assist one another during their collaboration, they established a ZPD (Gnadinger, 2008) between stronger and weaker students.

Gnadinger (2008) investigated the following two questions: (a) in what ways would peers provide scaffolding for one another during joint productive activities? (b) would children provide scaffolding, similar to that of adults, using the six means of assisted performances? The collection of data continued for 4 months using three resources: (a) videotaping, (b) informal interviews with teacher and students, and (c) field notes.

Analyzing the data revealed that students provided each other with three types of scaffolding: questioning, feedback, and instruction. The videotapes showed that 34 % of the scaffoldings were in the form of questions, and 21 % were in the form of providing instruction to one another. The study supported assertions about working in the ZPD

when peers would provide feedback to one another, and working together in the ZPD would help all learners and not only the weak ones.

Although Gnadinger's (2008) study reached the conclusion that peer mediation was effective, it mentioned that an adult collaborating with a child would lead to optimal learning. She concluded also that peer mediation was the best alternative while the teacher would be busy with a different small group of students. The current study benefited from this conclusion by asking participants to give each other gradual hints during their collaboration on the task-based activity and while the teacher-researcher was busy providing another group with the dynamic assessment hinting process.

One of the major limitations mentioned in Gnadinger's (2008) study was the lack of measurements. Therefore, it remains unknown to what extent students benefitted from their scaffolding. The absence of measurement was not only lacking in this study but also a common factor in all reviewed studies on dynamic assessment. Gnadinger's (2008) study was not conducted for a second-language classroom. Although this study was conducted in a classroom setting, it was still about peers helping one another and not about the teacher's role in the mediation of dynamic assessment. This study was about using students' collaboration to work on tasks for the purpose of prompting peer-mediation and the researcher finding that peer-mediation was good for all students regardless of their abilities was promising. Task-based language instruction was used in the current research to promote not only peer-mediation but also students-teacher mediation.

Anton (2009) conducted another dynamic-assessment study on advanced second language learners. Five third-year Spanish majors completed the entry exams announced

for the incoming students during the gathering of information for this study. The five exams included grammar, vocabulary, listening comprehension, reading comprehension, speaking, and writing. The five students reached or surpassed the minimum required scores. The speaking and the writing test followed the dynamic-assessment procedures to identify the abilities and weaknesses for each student. The speaking dynamic-assessment interviews were evaluated following the guidelines of the American Council on the Teaching of Foreign Languages scale for proficiency, and students went through four sections in these 10 to 15-minute interviews.

In the first section (2 minutes), the examiner asked the interviewee a few personalized questions about the examinee's personal interests, background, and past trips. The second section prompted the interviewee to narrate in the past about a picture provided; students needed to start their narration by saying yesterday. This section was conducted in three parts: (a) narrating without any help, (b) providing assistance and guidance by the examiner when necessary, and (c) the examiner narrating the story for the student to narrate it again. The third part was done only if needed. This second section of the speaking interview was designed to provide the interviewee with the DA scaffolding. In the third section, the examinee was asked to play the role of one character in the story to say something appropriate to the situation. Finally, the interviewee was asked to develop a 3-minute monologue on one of two topics. If the examinee was unable to sustain 3 minutes, the tester would guide the student with some further questions (Anton, 2009).

The score was based on what the student was able to do with the help provided, and in addition to this numeric score a qualitative report was provided. A qualitative

analysis of the results showed that “DA allows for a deeper and richer description of learners’ actual and emergent abilities, which enables programs to devise individualized instructional plans attuned to learners’ needs” (p. 576). Although these results were encouraging, there were certain aspects that were not addressed in this study. The most crucial of which was the remedial (instruction) program that was designed supposedly to improve the diagnosed problem. The research explained the interventionist approach of dynamic assessment only but did not elaborate on how the diagnosed weaknesses were addressed during the actual instruction program. Consequently, the activities used in the classroom to implement the dynamic-assessment gradual hints remained unknown.

Anton (2009) mentioned that the individualized attention of dynamic assessment would promote learning, regardless of its tediousness and time consumption that were the discouraging factors for teachers in the field to use dynamic assessment. The report was referring to the interactionist approach of dynamic assessment when describing the process as being labor intensive and needing long time in the last statement. No solutions were suggested or investigated for this issue. It would have been informative and satisfying, if the study included the use of the interactionist approach in a remedial program to follow the diagnostic exams. The reader of this report remained uncertain about the way the American Council on the Teaching of Foreign Languages guidelines were used in implementing dynamic assessment during these writing and speaking dynamic assessment instruments of the five diagnostic tests. Therefore, this current dissertation intended to elaborate on using the Interagency Language Roundtable as rubrics of identifying the linguistic abilities and weaknesses in a practical way. The

classroom-dynamic-assessment setting was investigated for this process to overcome the dynamic-assessment disadvantages mentioned above in this paragraph.

The following reviewed studies addressed dynamic assessment in a classroom setting. Poehner (2009) mentioned that this study conducted the dynamic-assessment approach on a group, and it focused on mediation in the second-language classroom. The author mentioned that classrooms did not permit the one-to-one mediation that characterized most DA work and ZPD (Poehner, 2009). The background information that emphasized the importance of social interaction as a medium for development was applied by Poehner (2009). He referred to Vygotsky's (1978) work describing humans, unlike other animals, interacting with the world in a mediated rather than a direct fashion. The report also mentioned that, according to Vygotsky (1978), human development would happen in two stages. The first stage is through "intermental plane" that take place during collaborative work with others and with cultural artifacts (p. 472). Then later, the development continues through the intramental functioning.

The study differentiated between the "group-as-context" and the "group-as-collective" (p. 474). According to this study, nothing would connect the members of the first group except time and space, and the group would represent only a "backdrop" to the performance and development of the individual. In the "group-as-cooperation," each individual would help the group interdependently to accomplish a goal that no one member can accomplish alone. The mediation would happen with both types in two ways "primary" and "secondary" (p. 477); the first would be when the teacher interacts with the learner directly, and the second would refer to the students benefiting from a primary interaction between the teacher and one of them. Two group-dynamic assessments would

develop as a result of these two ways of interaction. The first one would be the “concurrent group-dynamic assessment,” and the second would be “cumulative group-dynamic assessment” (p. 478).

The first one occurred when the teacher responded to a student who was facing a difficulty or having a question, but the whole group started participating immediately in a series of primary and secondary interactions. Cumulative group-dynamic assessment referred to the situation when the teacher conducted a series of one-on-one with different students while the group was collaborating toward one goal. The teacher-researcher of this present dissertation used both types of group-dynamic assessment. He used them during his responses to questions and needs from students while they were working in their small groups on their assigned tasks and while each group represented their product to the whole class. The teacher-researcher thought that students would benefit from both kinds of group-dynamic assessment while they provided each other with peer-mediated assistance as mentioned above in Gnadinger’s (2008) study. Poehner (2009) referred to a study that was still in press at the time of publishing his article (Lantolf & Poehner, 2011).

Lantolf and Poehner (2011) conducted a research of dynamic-assessment principles implemented in the context of a laboratory of a primary-school affiliated with a major urban university in the Northeastern United States. This school employed a second-language Spanish teacher with the pseudonym, Tracy. Tracy developed a unit around Peru that introduced students to a number of cultural topics and its relevant vocabulary. She prepared a cube that had a different animal on each side, and one student would volunteer at the time to go at the front of the class to roll it. This volunteer

described the animal while the other students watched. Tracy, who was also in front of the class, intervened to mediate when students had difficulties. She followed her interpretation of the dynamic assessment teacher's guide (Lantolf & Poehner, 2007) by providing the gradual hints as follow: (a) pausing, (b) repeating the whole phrase questioningly, (c) repeating just the part of the sentence with the error, (d) asking about what was wrong with that sentence, (e) pointing out the incorrect word, (f) asking either-or question, (g) identifying the correct answer, and finally (h) explaining why.

Tracy held a board and recorded the level of mediation that she provided for each one of six fourth-grade students. After using a table to show the way, Tracy recorded her hinting system on her board, Lantolf and Poehner (2011) reported three transcriptions of exchanges with students to demonstrate that the dynamic-assessment approach improved students' performance and how Tracy implemented the hinting process. The study concluded that group-dynamic assessment's contribution to second-language education was that it emphasized that classroom interactions were more systematic and more attuned to learners' developing abilities. In both the concurrent and cumulative formats, the teacher proceeded from a developmental perspective that informed her moment-to-moment assessments of the students' needs. On a different note, although the transcribed exchanges showed that students as secondary or primary "interactants" benefitted from the teacher's mediation, Lantolf and Poehner (2011) were not certain if the activity used was sufficient in keeping all students engaged. They were not sure whether all students were paying attention while the teacher was providing the gradual hints. They argued that organizing classroom activity in this way would enable teachers to explore and promote the group's ZPD while also supporting the development of learners individually.

To confirm and further their findings by using the Interagency Language Roundtable scale, the teacher-researcher of the present study considered their findings in designing the daily classroom's task-based instruction and in developing practical techniques for filling out the rubrics form of this study. While one of the groups was presenting, for example, others were tasked to critically express agreement, disagreement, or suggestions for improving the real-life solution presented. The teacher-researcher asked the other students in the presenter's group or in the whole class to respond to his provided hint first before he supplied a more explicit hint. The teacher-researcher started with Tracy's technique (Lantolf & Poehner, 2011) to fill out the rubrics form that he carried around on a clipboard.

Hill and Sabet (2009) conducted another study on dynamic assessment in a classroom setting, and this one was titled Dynamic Speaking Assessments. The study focused on four possible dynamic-speaking-assessment approaches. These four approaches were: the mediated assistance, transfer of learning, ZPD, and collaborative engagement. Mediated assistance took place between a teacher and a learner to identify a problem in the speaking performance. Transfer of learning evaluated the student's ability to transfer what they had learned initially to new situations. The learner's ZPD could be collective for a group of students who were solving a problem, and the focus here was on the sociocultural aspect of the ZPD. In this case, Hill and Sabet (2009) called it a group-ZPD, and that comparisons among individuals in this case were not as important as the dynamics of their cooperative activity became more relevant. The last dynamic-speaking-assessment approach was collaborative engagement, which diagnosed the problems occurring during the activities of dynamic speaking assessment. This year-long study

involved four speaking assessments of a first-year class at a university in Japan. Eighteen students participated in this study (12 female and 6 male). Each assessment involved two role-plays.

The first assessment was a nondynamic-assessment test that was used as a control. The second assessment was the first dynamic speaking assessment, and the second role-play of this assessment graduated in the transfer of learning difficulty. The third assessment used transfer-of-learning role-plays and mediated assessment in the form of recasts. In this test, the top one to nine students were paired with the lower 10 to 18 students, respectively. The final test used gradual transfer-of-learning role-plays, but mediated assistance was used at this time to evaluate the internalization of language features that were not previously demonstrated. The results of these dynamic-speaking-assessment approaches led to several conclusions.

The data suggested that transfer-of-learning role-plays of gradual difficulty were a genuine means of assessing the development of second language acquisition. Pairing and sequencing students by level were conducive to the interconnection of their performance, improving the dynamic-speaking-assessment performance, and reducing variation in their performance. Mediated-assistance data suggested that it had considerable cumulative improvement in dynamic speaking assessment and in the students' reciprocity in the form of recasts. Group-ZPD emphasized the fact that the ZPD was sociocultural. These positive findings left the reader wondering about the scale that was used to evaluate students' performance during the occurrences of mediated assistance, transfer of learning, collaborative engagement.

Hill and Sabet's (2009) study demonstrated that showing students needing lesser number of hints or eventually needing only implicit one reflected that internalization occurred for the language feature handled at that time. The kind of activities that prompted students to collaborate in class during the study was still unclear. The study did not mention the considerations made for crafting its classroom activities for the participating adult university students. The study did not justify the sufficiency of using the students' linguistic level as the only consideration for pairing students. Moreover, this study used pairing as the only grouping technique, and therefore left the reader wondering if this is the only suitable grouping approach for dynamic speaking assessment? If other group sizes were effective also, would the differences of adult profiles affect the collaboration efforts? Hill and Sabet's (2009) study failed to address this factor also.

The current study benefited from the four approaches of dynamic speaking assessment investigated in Hill and Sabet's (2009) study. The teacher-researcher of the present study designed task-based-language-instruction activities (please see Appendix E) that would permit him to provide mediated assistance to students in their small groups (pairs or groups of three students). These mediated assistance occurrences were in gradual standardized hints each of which became more explicit incrementally. These incrementally explicit hints helped the teacher-researcher of this study to diagnose the students' mature and immature abilities. He was able to identify how many increments of explicitness each student, a group of students, or the whole class was away from performing independently a particular language feature as described for the targeted

proficiency level. For this purpose, the teacher-researcher designed rubrics for this study that were based on the Interagency Language Roundtable scale.

This current study used transfer of learning as well. Transfer of learning was used in this study by observing students' performances while using the same language feature in different or more complex contexts. Using the same feature independently or with less explicit hints was indicative of the student's level of internalization of a certain feature. A student who performed independently a certain feature in several contexts while needing assistance for another emerging syntactical feature meant that he or she had become ready for another cycle of hints to learn the newly appearing erroneous utterance. The teacher-researcher provided these assistances to students while collaborating in their small groups to generate a measureable product (a solution that would need a language outcome). This context of task-based activities created the needed venues for Hill and Sabet's (2009) collaborative engagements. Collaborative engagements, mediated assistance, and transfer of learning were conducted in the students' ZPD. The teacher-researcher identified the students' ZPDs by conducting interventionist dynamic assessment interviews prior to the instruction phase of this study.

Brown (2009) conducted a study using dynamic assessment in a classroom setting to address a debate on foreign language instruction about the correlation between foreign language uptake and contact time. Brown (2009) reviewed the literature and found that studying advanced-level material would not guarantee uptake without the instructors playing an active role in negotiating the meaning of advanced-level forms of speech. Brown (2009) stated that this active role could be best done in the learner's ZPD in the form of scaffolding. The study showed that collaborative learning would be the key for

advanced learners' activities through content-based instruction, which would correspond very well to task-based language instruction (TBLI).

Based on these findings, Brown (2009) designed this study as a response to the growing demand for highly proficient speakers of foreign languages from both the private and Government sectors. The study intended to answer two questions: (a) Can sublevel or threshold oral and written proficiency gain be achieved in an advanced-level foreign language classroom over the period of an academic semester? (b) Does the application of debate activities carried out in the target language lead to measurable gain in oral and written proficiency? To answer these questions, 14 students in third- and fourth-year Russian classes at the beginning of the Fall 2006 semester were recruited to participate in this study (N. A. Brown, 2009). They were informed that a course would be offered in the Winter semester of 2007 titled Russian 49R: Global Diplomacy and Debate. Applicants were informed that those admitted would be eligible for a roundtrip travel to Russia as part of the course. During this trip, students would participate in a parliamentary style debates and Model United Nations competitions held at Russian State University for the Humanities in Moscow and at Saratov State University.

Pre- and Post-Oral Proficiency Interviews and Written Proficiency Tests were administered to 14 students selected to participate. Students received the guidelines of the American Council on the Teaching of Foreign Languages that constitute the standards for speakers and writers at the advanced level. Students alternated on different teams weekly during class sessions so that they do scaffolding to collaborate with or compete against each one in class at least once. The class met once a week for 2 hours, and they were assigned homework. Students were assigned weekly homework that was relevant to the

debate topics, and they were reading assignments in Russian and English. With two weeks remaining in the semester, students underwent a post-Oral Proficiency Interview and a Written Proficiency Test, and the findings suggested a general trend toward improved proficiency at .05 level of statistical significance for oral proficiency.

Brown (2009) reported that threshold gains exceeded sublevel gains in oral proficiency and written proficiency and “gainers” progressed incrementally with sublevels and across thresholds. Progress happened from Advanced-mid to Advance-high or Advance-high to superior. Brown (2009) expressed that one possibility for gaining proficiency could be related to Vygotsky’s (1978) idea of ZPD, because the level of difficulty used for the debates were within the ZPD of students who were rated Advanced-high in the pre-Oral Proficiency Interview. They were not within the ZPD of students who were rated Advanced-mid in the pre-Oral Proficiency Interview. Brown (2009) explained that this was the reason for three out of four students advancing from Advanced-high to superior, whereas only two out of seven students progressed from Advanced-mid to Advanced-high. Consequently, this current study made sure that the input material used was suitable and of interest to students by having students participate in selecting authentic passages of which the level of difficulty would be at “i+1.” The teacher-researcher would review their suggested material to make sure that its level of difficulty was suitable for all participants. However, the participating students did not suggest any material during the study’s course of instruction. The participants also were selected according to their proficiency level to eliminate any substantial difference among their proficiency levels as much as possible. This way, peer-scaffoldings could be more effective.

Other than scaffolding, Brown (2009) recorded that other factors needed consideration. Motivation was one of these considerations. For example, the background questionnaire showed possible correlation between the career interest of some students and the nature of the course curriculum and design, while it did not show the same for others. Brown's (2009) study actually referred to both the Interagency Language Roundtable (ILR) scale and the American Council on the Teaching of Foreign Languages guidelines (ACTFL), but the study failed to explain how either of the two scales was used to identify students' needs or progression. The limitation that remained unanswered was how the students' daily progress during the debates was tracked and used for subsequent lesson planning. The reader would wonder if the rotation of students on teams was sufficient in addressing their different learning styles. This dissertation attempted to use the Interagency Language Roundtable-based rubrics, text typology, task-based language instruction, a biographical background questionnaire, and identifying the students' intellectual styles to address these points.

Task-Based Language Instruction

The word task was used loosely in the field of Foreign Language Education for a long time and until experts found an acceptable definition for it (Ellis, 2009b; Nunan, 2004). Nunan (2004) criticized M. H. Long's (2000) task definition for being "nontechnical" and "nonlinguistic." He even described it by being the kind of response that he would obtain from a person in the street. Ellis (2009b) listed in a table several task definitions of other writers. These definitions either were not comprehensive enough for the pedagogical purpose and nature of a task as it is used in classrooms or they were not technical enough (Ellis, 2009b).

Ellis's (2009b) definition focused more on the process than on the end product of the activity, although one would assume that a work plan usually would include accumulative steps that lead to a deliverable product. Focusing on the process and not the product was evident in most sentences of Ellis's (2009b) definition through his explanation that the task's steps prompt learners to focus on meaning while collaborating on the end product. Ellis (2009b) advocated evaluating the content of the outcome holistically without evaluating its form or structure. The main goal was that the content of the proposed outcome is delivered correctly and appropriately. Appropriately in this context referred to the suitable fulfillment of the task at hand for its social context (Canale & Swain, 1980; Swain et al., 2010). Evaluating a task would be task-driven and not construct-driven (Anton, 2009; Bachman, 1990, 2002; Messick, 1994) when the focus was on the process more than being on the form or a product-focused. The conveyance of meaning through learner's incorporation of their own linguistic resources was the focus in this definition. The last statement demonstrated further that Ellis's (2009b) definition was process-driven; he chose to use the words "task can engage" to reflect the meaning of task as a venue in which learners could process receptive or productive skills.

Although Nunan (2004) agreed with Ellis's (2009b) definition for task-based language instruction, he still defined it in his words. Nunan's (2004) definition intended to emphasize all the stages of a task and not only the students' collaborative process but also emphasized from the very beginning the difference between a pedagogical task and regular daily tasks outside of the classroom (Bachman, 2002; Bachman & Palmer, 1996). Nunan (2004) defined the term task as a piece of work in a classroom. This classroom work was one piece, meaning an undivided block, with a beginning, middle, and an end.

Students' multimodal collaboration was expressed by Nunan's (2004) definition as involving learners' receptive and productive skills. There was a stronger emphasis on learners integrating both kinds of skills in Nunan's (2004) definition than how it was in Ellis's (2009b). Ellis (2009b) mentioned that a task could engage productive or receptive skills and not both as it was in Nunan's (2004). Moreover, Nunan (2004) mentioned grammar and the use of the target language openly, yet he explained that students would deploy their grammatical knowledge collaboratively for the purpose of conveying the meaning and generating the outcome. The main driver was the meaning conveyed during the process of collaboration to generate the task's end-product.

Although Ellis's (2009b) definition reflected more emphasis on the product than Nunan's (2004), the later reflected the importance of all the different steps of a task. All the different steps of task-based language instruction were important for the purpose of this current study, because both traits of task-based language instruction, instruction and assessment were used (Bachman & Palmer, 1996). Both the process and the product were important for the purpose of providing dynamic assessment to promote learning and to continually diagnosing students' needs.

Therefore, for this present study the following was the definition for task-based language instruction and it is similar to a definition found for task-based performance assessment (Bachman, 2002). A task would be a collaborative engagement of learners in small groups using their Arabic receptive and productive skills to think critically for the purpose of generating a measurable outcome relevant to real-life situations. The teacher's role would be to assist students during all stages from the beginning of the activity until they would deliver their product by providing calibrated and standardized gradual hints to

assess accurately their abilities and needs on the Interagency Language Roundtable scale. This definition was very close to task-based language performance assessment, which is a combination of task-based language instruction and performance-based assessment. Therefore, the following section explores the literature on performance-based assessment.

Performance-Based Assessment

Performance-based assessment is a broader term including other types of assessment, and these are authentic assessment and alternative assessment (Johnson, Penny, & Gordon, 2009). It is the type of assessment in which examinees were given the opportunity to demonstrate their knowledge and skills through their engagement in a process or delivering a product (Bachman, 2002; Foster & Skehan, 1996). This type of assessment should include four elements (Bachman, 2002): (a) a purpose, (b) tasks or prompts, (c) a response demand, and (d) a systematic method of rating performance. All these factors are in the definition of task-based language instruction as it was used in this study by combining task-based language instruction with dynamic assessment. In addition to these four factors of performance-based assessment, authentic assessment added the element of directly evaluating students' performance of a real-life situation that demanded their collaboration in small groups and critical thinking (Foster & Skehan, 1996). Evaluating performance in a real-life situation was fulfilled by the proficiency-based ILR scale used in this study. Moreover, authentic assessment would gather systematically information by evaluating every student's direct performance of a real-life task in the classroom over time and by using relevant rubrics that would be public and known to students for the purpose of giving students meaningful and accurate feedback

(Foster & Skehan, 1996). Therefore, the teacher-researcher gave a presentation to participants to make them aware of the rubrics used in this study.

Neither authentic assessment in particular nor performance-based assessment in general would be a replacement of traditional testing, but, in addition to test scores, it would gather more information daily about students' performance in class to compare their current abilities and inabilities against the requirements of real-life or end-of-course tasks (Anton, 2009; Bachman, 2002; Foster & Skehan, 1996). The information gathered systematically would include students' self-assessment and peer-assessment so that students were part of their own evaluation; these two types of assessments would prompt students to reflect on their own performance and compare it with the program objectives. Implementation of these two types of assessment was the main reason that the rubrics against which students were evaluated was known to them from the beginning to the end. Their awareness of the criteria would invite their involvement in developing classroom tasks that would help their learning process toward the end objectives of their class (Anton, 2009; Foster & Skehan, 1996).

In developing these rubrics, teachers would consider whether the standards were task-driven or construct-driven (Anton, 2009; Bachman, 2002; Messick, 1989, 1994). This consideration would drive the designing of the task, the prompt, to be suitable for either one of the two types. On the one hand, if rubrics were task-driven, then the evaluation would usually be holistic and criterion-referenced, that is, the rubrics were more product-driven. On the other hand, if the task were construct-driven, then the rubrics would tend to be more process driven, and the standards of the accuracy factors for performing the targeted construct would be included. The standards of accuracy

factors were in the core of the present study's rubrics. These two concepts of the evaluation being task-driven or construct-driven raised some concern, although the interest in both performance-based and authentic assessment was increasing in the field of assessment (Bachman, 2002).

Bachman (2002) expressed that task-based language performance assessment was prompting researchers and practitioners to reconsider many fundamental issues about what needs to be assessed. He explained his concern about the need for "evidence" and "representativeness" in evaluating the skills, knowledge, and ability of examinees exposed to either task-centered or construct-centered rubrics. One issue Bachman (2002) mentioned was the real-life tasks versus assessment tasks, and the possibility of the later to represent the complex and multiple possibilities of the earlier ones. His concern was that the few tasks used in the test could not represent enough all the possible array of tasks that people of the target language were expected to do in real-life situations. Another problem for him was the "difficulty with difficulty," meaning that people could do the same tasks at several hierarchical levels of performances. To solve these issues, he suggested the following to move forward: (a) conceptualizing tasks as sets of characteristics, (b) distinguishing among characteristics inherent in the task, the test-takers, and (c) the interactions between these two.

These suggestions were considered in the Defense Language Institute Foreign Language Center's Oral Proficiency Interview in particular and the Interagency Language Roundtable standards that were used in evaluating the speaking abilities of a foreign language. Many of the tasks required for the examinees' performance during the Oral Proficiency Interview could be considered a product and a process at the same time. For

example, tasks required for the speaking skill at Level 2 on the Interagency Language Roundtable scale were narration in the three time frames, describing a physical object, giving instructions, reporting facts on current events, and doing a role-play for a survival situation with a complication. All these tasks were a set of characteristics and require processing the language as much as they are products themselves (ILR, 2013a). Authentic assessment as described above is a great venue for students to carry out all these tasks during their collaboration to generate an outcome as described above (Foster & Skehan, 1996). All these tasks could be the product at the end of the students' collaboration as well.

Moreover, authentic assessment and task-based language instruction fulfill the principles of adult learning. Baron (1995) listed the following as advantages of authentic assessment: (a) authentic assessment techniques would measure directly what teachers want learners to know, (b) authentic assessment techniques would emphasize higher thinking skills, personal judgment, and collaboration, (c) authentic assessment would urge students to become active participants in the learning process, and (d) authentic assessment would allow and encourage educators to teach to the test without destroying validity. These advantages of authentic assessment would support the purpose of this study, which was to combine all the above mentioned variables and principles regarding adult learning by combining task-based language instruction with dynamic assessment.

Previous TBLI Studies

Considering the need for planning effective tasks for the current study, the following reviewed studies were selected. Skehan and Foster (1999) studied the influence of task structure and processing conditions on narrative retellings. They found that three

areas were competitive with one another over students' mental resources; these areas were fluency, accuracy, complexity or range, and, therefore, task research focused on three areas. These areas were (a) how balance might be achieved among these different three performance areas, (b) how task characteristics could influence performance and influence balance among the goals, and (c) how task conditions could influence performance and influence balance among the goals.

Three task types were used: personal tasks, narrative tasks, and decision-making tasks. Skehan and Foster (1999) reported that previous studies found that personal tasks would lead to higher fluency, narrative tasks would lead to better accuracy, and decision-making tasks would generate complex performances. Skehan and Foster (1999) found that pretask planning would lead to higher fluency, accuracy, and complexity in performing a task, although the authors thought that anticipating the topic and the kind of vocabulary of a task would be the reason behind these impressive results.

The purpose of the Skehan and Foster (1999) study was to explore how performance of a task could be affected by the degree of structure within the task and to explore how different processing conditions could influence performance. They hypothesized four conditions for tasks. The first two hypotheses were (a) a task with clear structure would lead to more fluent performance and (b) accuracy than tasks without such clarity. Third, a task structure would have no effect on the complexity of performance. Fourth, there would be an inverse relationship between the processing requirements of the task conditions and the accuracy, fluency, and complexity of the language generated. The research was designed to use two tasks and four performance conditions. One task was structured and the other task did not have predictable structure.

The four conditions came in different settings. Participants were to watch a video and to describe the story simultaneously, and the second setting was for participants to be told briefly the storyline before they watched the video and described the story simultaneously. In the third setting, participants were allowed to watch the video first and then had to describe the story as they watched it again. In the last setting, participants watched the video, and then they would retell the story in their own words. There were 47 participants (16 male, 31 female) from a wide variety of first-language backgrounds, studying English as a foreign language at Thames Valley University. They had a similar proficiency level in English according to the results of their placement test. They were assigned to six intermediate-level classes.

Over a few weeks, students were selected randomly by the class teacher to take part in the research. Then participants were assigned randomly into one of two tasks. The results after conducting an analysis of variance showed that the fluency of performance was found to be affected strongly by the degree of inherent task structure; more structured tasks generated more fluent language. In contrast, complexity of language was influenced by the processing load. In addition to these findings, Skehan and Foster (1999) concluded that accuracy of performance seemed dependent on the interaction between the task structure and the processing load.

In the same year, Foster and Skehan (1999) published a relevant research to the one mentioned above, and its title was *The Influence of Source of Planning and Focus of Planning on Task-Based Performance*. They wanted to see how performance would change, if the source of planning was teacher fronted, solitary by students, or group-based by students. Planning here did not refer to the teacher preparation and designing of the

task. Rather, it referred to the pretask activities in the classroom. The researchers wanted to investigate two foci in these pretask planning. The first was language and the second was the content.

This study was conducted on 66 students from six intermediate-level English classes at a college for adults (Foster & Skehan, 1999). Most of the students were in their 20s, and they came from a variety of first-language backgrounds. Only 13 of these 66 students were males. The researchers used a decision-making task under six conditions: (a) teacher fronted planning with focus on language, (b) group planning with focus on language, (c) teacher-led planning with focus on content, (d) group planning with focus on content, (e) solitary planning, and (f) no planning. Each one of these conditions was assigned randomly to one of the six classes. These conditions generated a 2x2 research design in which one dimension was the focus (language or content) and the other dimension was source (teacher or group).

Considering that the underlying rationale was that there was limited-capacity for processing ability in which concern to be fluent, to be conservatively accurate, and to take risk producing complex language need to be balanced. Foster and Skehan (1999) concluded, first, that teacher-led planning produced the highest levels of accuracy, and led to a greater avoidance of error. As for complexity, “the implication here was that there was a role for the teacher in pretask work, to channel attention and to ensure that the language used in the task would make a pedagogic contribution” (p. 238). Second, the results for group-based planning were not positive, because “it appears that student groups do not operate as efficiently as when either the pretask preparation time is organized by the teacher, or when learners are able to work independently” (p. 238).

Third, instructions that focused on language and on content did not produce different results.

With these findings, Foster and Skehan (1999) provided integrated assessments of the four different source-of-planning conditions. Solitary planning would generate greater complexity, and the learner would become able to take a long turn in discourse. The teacher-fronted planning would generate clearer accuracy effect, and it would lead to more control over the language. This accuracy was not on the expense of fluency, which was pleasingly surprising. The group-based planning proved to be an unsuccessful condition in this study, and the results were undistinguishable from the comparison group. The comparison group generated less complex language as it had been the case in previous studies.

Ellis (2009a) reviewed previous studies in a research on the differential effects of three types of task planning on the fluency, complexity, and accuracy in second-language oral production. Planning in this context referred to the time given to students to prepare for the task. Ellis (2009a) referred to his own categorization of task planning (Ellis, 2005) that sorted it into three types. These three types of planning were rehearsal, pretask, and within-task. Actually, there was a basic distinction between pretask and within-task planning. The first could be divided into rehearsal and strategic planning, and the second was sorted into pressured and unpressured in which the pressure referred to the availability of time for students to finish. The strategic planning had been the most researched in the field, and Ellis (2009a) reviewed 19 studies and synthesized them into a comparison table. Strategic planning meant that students prepare the content of the task

product without rehearsing its delivery (Ellis, 2009a). Then, Ellis (2009a) summarized his extensive literature review of the previous task studies.

Rehearsal would lead to greater fluency and complexity; the improvement on accuracy was to a lesser extent. These effects did not transfer to new tasks unless an intervention was used (Ellis, 2009a). Dynamic assessment was the type of intervention that this study intended to use. As far as strategic planning, it would benefit fluency, but the results were mixed where complexity and accuracy were concerned. Ellis (2009a) explained the reason for this inconsistency by the learners' limited processing capacity (Skehan, 1996). Some variables had an effect on strategic planning, the clearer of which was the learners' proficiency level. Strategic planning was less evident in very-advanced learners as much as it was in beginners. Planning was of greater benefit for less "well-structured tasks." Finally, within-task planning might benefit complexity and accuracy, but it would not have a detrimental effect on fluency.

Summary

Dynamic assessment (DA) was based on Vygotsky's (1978) ZPD and was used for diagnosing students' weaknesses and strengths and for promoting the acquisition of language. It was used usually for diagnosing purposes at the beginning of a language program, and this approach was known as interventionist DA (Grigorenko & Sternberg, 2002; Poehner, 2005). When the purpose was mainly promoting language acquisition during regular instruction, the approach was known as interactionist DA (Grigorenko & Sternberg, 2002; Poehner, 2005). Most of the DA studies were conducted in a tutoring format; the few studies conducted in a classroom format showed positive results. Previous studies revealed the presence of a group-ZPD that could be used in classroom

activities to deploy two different classroom DA techniques: (a) cumulative DA, and (b) concurrent DA. The DA studies used in a classroom setting were based mainly on one-on-one interactions of peers. Previous studies did not show the daily language acquisition or diagnosis using the rubrics of a standardized scale. Previous studies did not elaborate also on the activities used in class with adult students (Anton, 2009; Brown, 2009; Gnadinger, 2008; Hill & Sabet, 2009; Lantolf & Poehner, 2011; Poehner, 2005, 2009).

The literature reviewed in this chapter included several definitions for task-based language instruction. All the definitions emphasized the importance of student-centeredness, simulating real-life situations, and creating a measurable product (outcome). The main point was that a task would require the collaboration of all group members to generate an original product. Foster and Skehan (1999), Skehan and Foster (1999), and Ellis (2009a, 2009b) conducted studies on the influence of pretask source planning and task structure on task-based performance. Group-based pretask planning did not lead to impressive findings in most studies, and therefore would be avoided in this study (Foster & Skehan, 1999). Teacher-fronted and student-solitary planning were conducive to accurate or complex performances. Foster and Skehan's (1999) point was that the teacher's role was crucial in the pretask planning to channel attention and to ensure that language used in the task would make a pedagogic contribution (Foster & Skehan, 1999).

As far as task structure and processing conditions on narrative retelling, Foster and Skehan (1999) reviewed the previous studies for three types of tasks: (a) personal, (b) narrative, and (c) decision-making tasks. These studies found that personal and narrative tasks would lead to higher fluency, whereas decision-making tasks would generate

complex performances. Skehan and Foster (1999) studied four conditions for tasks, and they found that fluency of performance was affected greatly by the degree of task structure, that is, more structured tasks generated more fluent language. They found the processing load affected the complexity of language. They concluded that the accuracy of performance was dependent on the interaction between task structure and the processing load. A solution for the needed balance was found in the literature. Rehearsal would lead to greater fluency and complexity, although the improvement on accuracy would be to a lesser extent (Ellis, 2009a). Ellis (2009b) reported that these effects would not transfer to new tasks unless an intervention was used. In this present study, the intervention would be introduced through the DA process in general and the transfer-of-learning reviewed at the beginning of this chapter.

Cumulative and concurrent group-DAs were two successful classroom-settings that this study generated by using task-based activities. In this situation, students were given sufficient time for pretask planning; students solitary or teacher-fronted planning were used in the current study (Ellis, 2009a; Foster & Skehan, 1999), and when the pretask rehearsal was used, it was combined with the DA-scaffolding assistance to promote language acquisition (Ellis, 2009a).

Crafting the tasks for this study considered the principles of adult learning (Dean, 2004; H. B. Long, 2004). They were relevant to real-life practical use and prompted students to use their collective critical thinking. These adult students had autonomy while generating their own unique products. Distributing learners on small groups in class was based on the types of their intellectual styles, biographical background, and on the nature of the assigned task-based activities (Zhang & Sternberg, 2005; Zhang, Sternberg, &

Rayner, 2012). Not only their biographical background and personal interest were considered for their distribution on small groups, but also, they were considered in selecting suitable material for their existing proficiency level. The input material was selected according to the speaking functions that students needed to perform for the targeted proficiency level.

CHAPTER III

METHODOLOGY

This chapter focuses on the methodology that was used in this dissertation, and it is divided into 10 sections. These 10 sections are: (a) research design, (b) participants, (c) protection of human subjects, (d) instruction, (e) instrumentation, (f) use of assessment, (g) background of teacher-researcher, (h) research questions, (i) data analysis, and (j) limitations.

The purpose of this study was to investigate the effectiveness of combining dynamic assessment with task-based activities that would target the speaking skill of Arabic (Goos et al., 2002; H. B. Long, 2004); task-based-language-instruction (TBLI) activities included small-group collaborations in Arabic for the purpose of creating measurable products. More specifically, this dissertation explored the effect of using an ongoing classroom assessment (Anton, 2009; Bachman, 1990) to gauge and exploit Vygotsky's zone for proximal development (ZPD) of each learner or a group of adult students of Arabic (Allal & Pelgrims Ducrey, 2000; Dean, 2004). Providing instruction through gauging and scaffolding into learners' ZPD was known in the field of foreign language education as dynamic assessment (DA). This mixed-method study was designed to contribute to the knowledgebase developed from previous studies of the effectiveness of DA-based instruction.

It investigated the practicality of continually assessing students' weaknesses and strengths during their course of instruction and particularly as a group (Brown, 2009; Ellis, 2009a). This research used the proficiency scale employed in the U.S. Government with students attending the Defense Language Institute Foreign Language Center. These

students were military service men and women, and they were learning Arabic in the institute's Basic Course. Therefore, the findings of how effective dynamic assessment would be in their daily classroom instruction might benefit language-adult-learning programs at DLIFLC, colleges, and universities around the world.

This dissertation investigated the effect of combining task-based-language-instruction activities in classrooms with dynamic assessment on the students' Arabic speaking abilities. The process of combining both of these approaches was referred to as DA/TBLI instruction in this study. The process of DA/TBLI instruction was guided and measured by the U.S. Government's proficiency scale known as the Interagency Language Roundtable scale (ILR). The study used Interagency-Language-Roundtable-based rubrics guided by a table format found in performance-based assessment (Johnson, Penny, & Gordon, 2009). The standards for the different targeted independent performances for students were established by deconstructing the Interagency-Language-Roundtable scale into recognizable sublevels for the ranges between the descriptions of every two existing proficiency levels (ILR, 2013a).

These recognizable sublevels enabled this study to accomplish its purpose, because they provided a valid and reliable measuring instrument for gauging the effect of dynamic-assessment-based instruction on both language learning and diagnosing students' needs. The study's rubrics measured the effect of dynamic assessment on language learning and on the daily diagnosing ability for students' needs. The Defense Language Institute Foreign Language Center had been using task-based language instruction in its language-teaching programs since 2003 and a process called Diagnostic Assessment since 1998. The Arabic schools used mainly Diagnostic Assessment two

times during its Arabic Basic Course. The daily process for diagnosing students' needs was accomplished mostly by the teacher's personal observation or by conducting Oral Proficiency Interviews during the program's formative-assessment system. The Arabic schools of DLIFLC offered students periodic Oral Proficiency Interviews toward the end of the basic course and prior to their formal exit test to provide them with diagnostic feedback.

The effect of DA/TBLI could be measured by comparing the change in students' performance using the Interagency-Language-Roundtable rubrics. Comparing the Oral Proficiency Interview to both types of dynamic assessment would illustrate their differences in evaluating Arabic in general and their diagnostic feedback in particular. To make this measuring more practical for the purpose of this study, the focus was only on one accuracy factor for each proficiency level of the Interagency Language Roundtable scale. The accuracy factor measured in this study was the "structural control."

Research Design

This mixed-method study (Creswell, 2007) was conducted in three phases: (a) Pre-DA, (b) DA, and (c) Post-DA. In the Pre-DA phase, each student's oral proficiency level was evaluated by an Oral Proficiency Interview (OPI) conducted by two certified testers. Students' diagnostic strengths and weaknesses were evaluated through interventionist-dynamic-assessment interviews in one-on-one sessions. The teacher-researcher not only conducted the interventionist interviews but also trained both students and testers in the dynamic assessment approach. The purpose of training the participating students was to make certain that they were familiar with the hinting process, the targeted descriptors of the Interagency Language Roundtable, and the dynamic-assessment logic

in helping their Arabic acquisition. Familiarizing students with the scale by which they were evaluated was compliant with the principles of authentic assessment as reviewed in the last chapter. As far as the participating certified testers, they observed the teacher-researcher's lessons during the DA phase and filled out the Dynamic-Assessment Rubrics Form (Appendix D) that was designed for this study. Therefore, they needed to have a clear understanding of the process and a common perception of the targeted descriptors.

In the DA phase, students attended one-hour lessons daily for 4 weeks in the third semester of their language program, which was the last semester of the Arabic Basic Course. These one-hour lessons were during one of their regular 7-hour daily classes. During these sessions, the teacher-researcher used the interactionist-dynamic-assessment approach to give feedback to students individually or to small groups while working on or delivering the measurable product of a task-based-language-instruction activity. The teacher-researcher used the Dynamic-Assessment Rubrics Form to diagnose and record the students' daily classroom performances (Appendix C and Appendix D). The teacher-researcher interviewed the observer immediately in a post-lesson session to obtain his or her feedback perception on the DA/TBLI instruction.

The following is an example for a possible dynamic-assessment interaction during which the teacher is providing the following gradual hints: (a) not accepting the answer, (b) referring to the accuracy factor, (c) asking questions, (d) repeating the specific erroneous utterance, and (e) providing the student with the correct answer and its explanation. The example is an English translation for a similar dialogue in Arabic.

Student: Last weekend ... I went with my friends to Los Angeles. After they arrived to the hotel and they put our bags in our room, we went to Disney Land.

(a) Teacher: not accepting the utterance by shaking his head questioningly with a gentle smile.

Student: Silently reflecting in confusion.

(b) Teacher: Syntactical control

Student: Still confused.

(c) Teacher: Who arrived to the hotel, and who put the bags in the room? Did you stay in one room?

Student: Oh ... after we arrived to the hotel and we put the bags in our rooms, they went to Disney Land.

(d) Teacher: They went to Disney Land?

Student: Oh ... we went to Disney Land

In this example, the teacher did not need to clarify or explain any grammatical feature for the student. If the student was not able to produce the proper utterance, the teacher would explain the plural conjugation of past tense verbs. In similar interactions, the teacher-researcher would enter number four in the appropriate box on the Dynamic Assessment Rubrics Form (Appendix C or D) to reflect the number of hints provided for the student. The teacher could write in the remarks section that this student or group was able to conjugate some of the verbs correctly. Notes for the teacher-researcher's reflections were entered in the teacher journal after each session.

In the Post-DA phase, students' proficiency levels were reevaluated by Oral Proficiency Interviews and dynamic-assessment interviews. The Post-DA Oral Proficiency Interview was conducted by two different testers and the teacher-researcher administered a final interventionist DA for each participant. Students were interviewed to

evaluate their perception of the DA approach. Interviews for both students and testers were recorded and transcribed. Transcripts were reviewed and coded for emerging themes. These themes were then analyzed in relation to student perceptions of the DA process. Additionally, students responded to a survey of ten 5-point scales to measure their perception of the DA/TBLI instruction. Numbers from one to five on each scale correspond to the following qualitative values: (1) strongly disagree, (2) disagree, (3) I do not mind/similar to regular instruction, (4) agree, and (5) strongly agree.

The following flowchart (Figure 1) shows the procedures of each one of the three phases of this study's research design as explained above in this section.

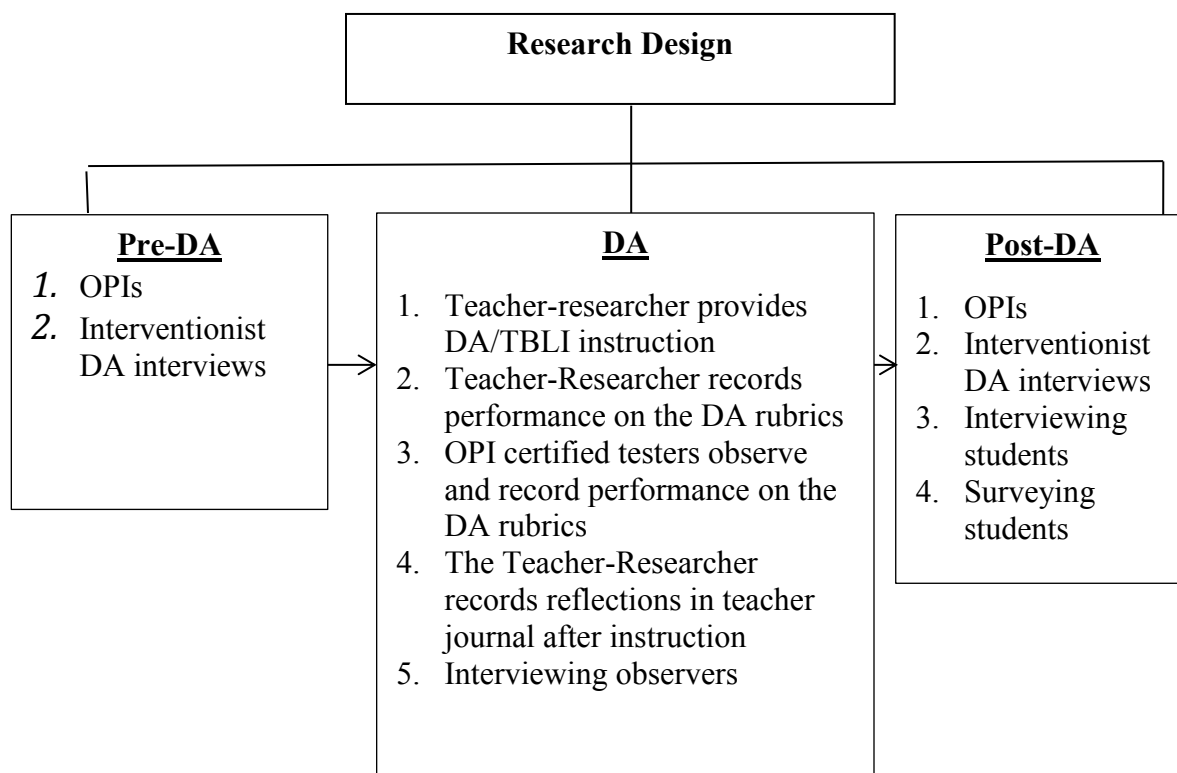


Figure 1 Research Design

Participants

The Defense Language Institute Foreign Language Center (DLIFLC) uses teaching teams and each has an average of six teachers. These teaching teams place every six students in one classroom during the Basic Course. Students of one of the classes attending the third and last semester in the Arabic Basic Course at the Defense Language Institute Foreign Language Center (DLIFLC) were recruited to participate in this study. Six out of the 10 volunteers who went through the pre-Oral Proficiency Interview (pre-OPI) to participate in this study were selected. The selection of the six students was based on the compatibility of their results of the pre-OPIs, intellectual styles, and biographical data. These students were referred to by their aliases to protect their confidentiality. These aliases are Basem, Hazem, Ibrahim, Jamal, Ramzy, and Salwa. The first five names were for the male students, and the last one was for the only female student in this group. These aliases were Arabic names given to students during their attendance of the Arabic Basic Course.

These students were attending the last 8 weeks of their 63-week long training during the DA phase, and their proficiency level was about “1+” on the ILR scale. Their proficiency level in both skills of reading and listening were assessed by conducting a recall-protocol periodically in class and by reviewing their last results on the regular recall-protocol conducted by the institute for its diagnostic assessment purposes. Students in the Arabic Basic Course attend classes for 7 hours daily, and this research was done during one of those hours. To identify the participants’ personal profile differences, they answered few questions on their background during the Pre-DA phase (Appendix B).

These questions were designed to identify each participant age, gender, military rank, social status (married, single, or having children), educational background, travels, previous work experiences, personal interests and hobbies (Appendix B). Knowing this information helped the teacher-researcher in selecting suitable and interesting material for the daily classes. The teacher-researcher obtained each student's profile, which included their grade point averages (GPA) at that time in the course in all modes by the institute's formative evaluation system (listening, reading, and speaking), their previous counseling statements by all teachers, and the initial assessment of their learning styles conducted prior to the beginning of their Arabic Basic Course.

Knowing this information about the participants helped in designing the classroom activities, selecting supplementary materials, and in dividing them into small groups of two to three students during classroom activities. Biographical data were not the only differences among adult students that were needed in designing classroom activities. Their intellectual styles were very important in designing classroom activities. Therefore, students answered the Thinking Style Inventory (Sternberg, Wagner, & Zhang, 2007) and the Myers-Briggs Types Indicator (Briggs & Myers, 1998) questionnaires during the Pre-DA phase, and the teacher-researcher evaluated his own intellectual styles at the same time as well. Empirical evidence found a positive correlation between the students' academic progress and having a teacher whose intellectual styles matching theirs (Fan & He, 2012).

If the teacher-researcher's intellectual styles did not match any student, it would not be a formidable problem. The reason was that intellectual styles are modifiable, because of being "states" and not "traits" (Zhang & Sternberg, 2005). Therefore, the main

purpose of evaluating the intellectual styles of the teacher-researcher and the students was to make each one of them aware of his or her own inclinations and preferences so that they would deploy the opposite construct when needed. There is empirical evidence in the literature showing a positive correlation between the students' awareness of their intellectual styles and their academic progress (Fan & He, 2012). Consequently, the teacher-researcher provided the intellectual style results with the participating students. The following section was designed for the description of each one of the participants as collected from the previously mentioned questionnaires. The Arabic aliases that were used in the classroom for these students will be referred to in the remaining part of this dissertation.

Basem is in the United States Marine Corps (USMC), and he was born in 1992. His wife also is serving in the USMC studying Tagalog at the Defense Language Institute Foreign Language Center. He joined the Marines immediately after graduating high school where he studied Spanish for 4 years. He loved reading, writing, video games, talking about history, myths, and religion. He lived in Australia for 5 years but also has been to Mexico, Fiji Islands, Canada, the Caribbean Islands, and many states in the US. His responses to the questionnaires for the Myer-Briggs Type Indicator (MBTI) and the Thinking Styles Inventory (STI) reflected an introverted learner who would prefer to focus on the present and concrete information. His answers reflected that he would learn better in well-structured activities.

Hazem is a male sailor born in 1990, and he attended college for one year before joining the Navy. He had learned French before he joined the military service. He likes to play video games, program computers, play soccer, and going out with his girlfriend. He

had traveled to France both as a student and as tourist, and he had been to Morocco for one month with the DLIFLC for the immersion language training. His responses to the MBTI reflected an introverted learner who would focus on the future, yet he would prefer structured activities.

Ibrahim is a Specialist in the Army who was born in 1986. He is married to a house maker who is an elementary-school teacher. He obtained a Bachelor's in Psychology and Religious Studies from the University of South Florida. He learned Spanish and Ancient Greek. He enjoyed computers, electronics, theology, video games, Poker, Mixed Martial Arts, and music. He traveled to Morocco with DLIFLC for a month but had been to the Caribbean Islands on a cruise with the family. He travelled all over the United States because his mother had various flight benefits. His answers to the MBTI and STI reflected an introverted learner who preferred concrete information. His answers, however, indicated that he would be flexible with the options available.

Jamal is an Army Major who has a Bachelor Degree in Computer Sciences, and he was born in 1979. He speaks French, because he attended the International French-American School for 4 years. Additionally, he has limited capabilities in Spanish. Jamal travelled extensively to include a 2-week trip as an exchange student to Tahiti where most of his teachers were French and could not speak English. He enjoyed very much the history of the Middle East, including the contemporary and ancient conflicts and developments. He also enjoys the application of technology, space physics, and computer applications. His answers to the MBTI and STI reflected an introverted learner who would prefer to focus on the present task and to base his decisions on logic. His answers also indicated that he would prefer to work in a structured environment.

Salwa is the only female Army soldier in this class. She was born in 1987, and she had a Bachelor Degree in Biology. She also studied Art History and Studio Art. She had learned Portuguese by attending classes and by living in the country for 3 years. She also had classes in Latin. She visited Spain and Rome while living in Portugal, and she has been to most states on both coasts of the United States. She liked reading, science, art, and outdoor activities. She enjoyed topics related to science in general and medical science in particular. Her answers to the MBTI and STI reflected an extroverted learner who would tend to focus on the present and concrete information. Her answers also indicated that she would fit in with any group of people, yet would perform better in a structured environment. She also showed that she was a very analytical learner who would focus on the details to understand the bigger picture.

Ramzy is serving in the USMC and was born in 1989. His fiancé was working as high-school history teacher in Louisiana. He studied Mechanical Engineering at Georgia Tech for two years. He had developed some limited abilities in French and Spanish before joining the Navy. He likes to read and play soccer and enjoys the physical work in the military training. He had a 2-week tour in England and Scotland with high-school friends. He likes topics that would give insights into the Middle Eastern cultures. His answers to the MBTI and STI showed him as an introverted learner who would tend to focus on the future and on what could be accomplished by following observed patterns. His answers reflected a person who would base his decisions on logic and would perform best in a structured environment.

Although five out of the six participants were shown to be introverted learners who would tend to be calm, their responses on the STI reflected great flexibility. This

flexibility of changing one's tendency was referred to as Type III Intellectual Styles by Zhang and Sternberg (2005). This flexibility coupled with their diverse background and interests were considered in selecting the passages for the daily lessons and in dividing these students into pairs or small groups to work on the assigned tasks.

Protection of Human Subjects

A prior approval by the Institute Review Boards of the University of San Francisco and the Defense Language Institute Foreign Language Center had been obtained through the proper and satisfactorily procedures for both organizations. The participating students signed an informed-consent form (Appendices A) before the beginning of this study, and this form mentioned their right to drop out of the study at any time. Although participating in the study did not grant them any financial award, they were promised quality instruction that would help them meeting the end-objective of their Arabic Basic Course.

All students available who were attending the third semester of the Arabic Basic Course were invited to participate in this study regardless of their gender, country of origin, military rank, faith, race, ethnicity, political affiliation, or any other personal background. Students were selected according to their proficiency level, intellectual styles, biographical background, and their diagnostic feedback for reading and listening. The teacher-researcher asked both students and certified testers to volunteer after giving a presentation explaining to them the dynamic-assessment approach and its expected process in the classroom and all the different steps of this study. The identity of involved students and testers were referred to by using aliases in this study as would be the case in any future publication.

Instruction

In the DA phase, the teacher-researcher used the Dynamic Assessment Rubrics Form (Appendix C and D) to record the effect of the interactionist dynamic assessment on the daily classroom's task-based-language-instruction activities (Poehner, 2005). Showing the students as needing fewer implicit hints than using more explicit hints was reflective of their improvement. The researcher used the interactionist dynamic assessment daily with the participants for one hour during which they collaborated on real-life tasks.

Pretask (Foster & Skehan, 1996) activities were conducted for 10 minutes before the task started. In these 10 minutes, the teacher used either one or a combination of solitary and teacher-fronted planning (Ellis, 2009a; Foster & Skehan, 1999) to obtain the best fluency and complexity possible during the execution of the day's task. Both fluency and language complexity as accuracy was not required to the same extent, were key factors in the descriptors of Level 2 for the speaking mode on the Interagency Language Roundtable scale. Level 2 for the speaking mode was the next measurable level by the Oral Proficiency Interview for the participants whose proficiency level at that point in the program should be at Level 1+ as mentioned earlier in this chapter. The day's task was done in small groups of two to three students to set the stage for the students' collaborative work. Students started working in pairs and then in two groups of three students or vice versa. The purpose of what was creating information, reasoning, or opinion gaps so that students interact critically in Arabic to generate a meaningful product.

Using the information gathered about each student's background, intellectual styles, and linguistic (listening, reading, and speaking) weaknesses and strengths relevant to the day's task, the teacher-researcher divided the students into small groups during the task-based-language-instruction activities effectively and to tailor the classroom material suitably to their interests or Level-2 functions. For example, knowing their biographic background helped in knowing the strengths and weaknesses in their knowledgebase to consider in the lesson planning process and in mixing and matching them into their working groups. The teacher-researcher knew their topics of interest from the biographical data, and he continued soliciting the students' opinion on the topics of the classroom input material for the whole duration of this study (Galbraith, 2004b). This action was guided by the principles of adult learning (H. B. Long, 2004).

The teacher-researcher used material suitable for the student's current proficiency level and his or her daily identified weaknesses as compared with the descriptors of the Dynamic Assessment Rubrics Form (Appendices C and D). The teacher-researcher used the principles of text typology (Child, 1987, 1998, 2001) to select any written or auditory authentic text as an input material, because text typology would describe the texts' different levels of difficulty. These descriptions were congruent with the ILR proficiency levels of people's abilities in using a foreign language. The teacher-researcher considered that the material used would follow the "i+1" formula as presented in chapter I (Krashen, 1982). Students were prompted to use the content of the authentic material to work cooperatively. Students used information, reasoning, opinion gap, or a combination of thereof to cooperate and present orally a measurable product.

During their collaborative work, students provided each other with the gradual hints as they were trained to do during the Pre-DA phase; this process of peer-to-peer DA was not recorded. It was incorporated in the small-group activities and in the students' final presentation for the purpose of promoting language acquisition. The second purpose was to create a ZPD between the teacher-researcher and the collective mind of a small group or the whole class. Once they asked for the teacher-researcher's help to overcome a difficulty, it meant that their group-DA was at that point where their aggregate knowledgebase was insufficient for the task at hand, and the teacher-researcher negotiated their group-ZPD through the established standardized hints on the Dynamic Assessment Rubrics Form (Appendix D).

Then, while students were still busy working on their assigned task, the teacher-researcher recorded the number that reflected the level of hinting in the suitable box on the Dynamic Assessment Rubrics Form (Appendix C). The teacher-researcher continued to provide his assistance using the dynamic-assessment approach until the groups were finished preparing and ready to present their final product. During the presentation of each group for their product, the teacher-researcher used dynamic assessment suitably. The word suitably meant that the teacher-researcher used dynamic assessment wisely to avoid lowering the students' fluency for the sake of accuracy. Recording the assistance provided was completed quickly by entering a number in the proper box on the form (Appendix C or D). At the same time, the certified-tester observing the class used the Dynamic Assessment Rubrics Form (DARF) to record the same process. The observer did not participate in the teaching process. He or she used DARF to record his or her understanding of the teacher-researcher's feedback to students, and the observer took

notes also while observing the classroom activities. Then, the observer discussed these notes and his or her entries on the Dynamic Assessment Rubrics Form (Appendix D) with the teacher-researcher immediately after the lesson ended. The teacher-researcher interviewed the observer to obtain his or her feedback perception on the DA/TBLI approach during this same meeting.

The Defense Language Institute Foreign Language Center had not used dynamic assessment in its Arabic classrooms, and the Dynamic Assessment Rubrics Form (Appendices C and D) were devised to explore the effect of dynamic assessment on the students' daily progress in speaking. The previous studies and literature did not specify a particular scale and consequently standards for what they considered an endpoint for the targeted independent performance (ACTFL, 2012; Alderson, 2005; Anton, 2009; Doolittle, 1997; Havnes, 2008; Lantolf & Poehner, 2011; Poehner & Lantolf, 2005). Although the Dynamic Assessment Rubrics Form was designed for recording and tracking students' performance in all the accuracy factors of the ILR scale (lexical control, structural control, sociolinguistic control, delivery, and text type which was the length of utterance), the teacher-researcher recorded only the structural control part to make sure that analyzing the data was practical for this study.

Instrumentation

This section presents the rubric used in both the interventionist DA and in the DA/TBLI instruction. The section discusses its validity and reliability and validation process. Following the presentation of the used rubric, the section presents the questions used in interviewing the students and the observers and the students' survey. The next part discusses the use of dynamic assessment. This ongoing classroom assessment

(Angelo & Cross, 1993) was criterion referenced (Bachman, 1990) and was developed by deconstructing the ILR standards for the targeted proficiency range on the Interagency Language Roundtable scale. This form used the Interagency Language Roundtable range between a high point into Level 1's abilities to a high point into Level 2 upward toward the descriptors of Level 3 on the scale (Appendix D). For example, the standards listed in the third column from the left were lifted faithfully from the Interagency Language Roundtable descriptors of speaking at Level 2 (DLIFLC, 2010; ILR, 2013a). The two boxes to its left reflected two weaker performances, and the two boxes to its right reflected two stronger performances in the range between Level 2 and Level 2+. In every row of the form, each sublevel used had a box underneath it for the teacher to enter the number that reflected the times of assistances provided and consequently their level of explicitness. The number for the hints provided reflected the level of explicitness needed for the learner to perform at the desired endpoint described in the box above it. Ultimately, the standards for the desired performance were those listed in the box all-the-way-to-the-right side of the form. The same format of the row described above was repeated in the rows below it to record the following attempts by the same student or group.

These subsequent attempts could show easily students' progress by comparing them with the level of assistance provided previously on the Dynamic Assessment Rubrics Form. The Interagency Language Roundtable standards were evaluated against the students' performance of the tasks for each proficiency level. The same tasks used by the institute's Oral Proficiency Interview, and DLIFLC conducts about 7,000 OPIs a year (DLIFLC, 2013b), which reflects that this test had matured to a very practical test with

high validity and reliability over the past 40 years (Child, Clifford, & Lowe, 1993). Using the same tasks or functions of the Oral Proficiency Interview and the ILR standards to conduct dynamic assessment in classrooms had not been used, and no other study in the literature reflected the using of particular rubrics based on any known scale (Alderson, 2005; Anton, 2009; Doolittle, 1995; Grigorenko & Sternberg, 2002; Havnes, 2008; Lantolf & Poehner, 2011; Poehner, 2005).

The Dynamic Assessment Rubric Form

The Dynamic Assessment Rubric Form (Appendix D) shows only the sublevels used in this study for evaluating the structural control during the students' performances. The structural control was one of five accuracy factors that normally would be evaluated for every proficiency level on the Interagency Language Roundtable scale during an Oral Proficiency Interview (DLIFLC, 2010). Students needed to be working on a task that would require them to be immersed in a simulated real-life situation to use their critical thinking for the purpose of developing a product. While the reporter of a small group presented their product or while the students of a group asked for assistance, the teacher-researcher or the observer circled the box that reflected the student's or group's representative performance, and used the box below it to enter the number reflecting the level of hinting. Hints ranged from level 1 that reflected that the teacher-researcher did not accept the answer to level 5 that meant providing the student with the answer along with its explanation. In between these two ends, the following gradual levels of assistance were provided: level 2 meant repeating broadly the erroneous utterance, level 3 indicated that the teacher-researcher repeated the specific erroneous utterance, and level 4 reflected naming the syntactical deficiency.

Validity and Reliability

Rubrics are pivotal to have a valid and reliable assessment (Bachman, 1990, 2002; Brown & Abeywickrama, 2010; Johnson et al., 2009). The teacher-researcher started the ILR-related part of these rubrics, in the past, by deconstructing the ILR scale to conduct formative-speaking assessment for students attending the Arabic Basic Course at the Defense Language Institute Foreign Language Center. The deconstruction of the Interagency Language Roundtable scale reflected the students' progress in between every two of its proficiency levels; every two ILR proficiency levels included five progressive descriptions that reflected gradual improvement. These deconstructed-ILR standards were merged into a table that included designated boxes for entering the level of the DA gradual standardized hints. This table allowed the teacher-researcher to record his assessment for the students' performance and the assistance provided to aid their demonstrated abilities. These rubrics had been validated through classroom activities, getting feedback from authorities, and using them for over 2 years in the formative assessment needed in the Arabic Basic Course.

This form was used easily in classrooms in which task-based language instruction was given and where students were collaborating to solve a problem in real-life situations. Their progression on the ILR scale while collaborating or delivering the products of the different tasks in class were measured by the deconstructed sublevels shown in Appendix D. Considering that the ILR scale was the one used in the Oral Proficiency Interview, the DA rubrics devised for this study gained consequently a high level of validity. This high validity was due not only to measuring students' progression by using the same criteria of the Oral Proficiency Interview but also because the ILR

tasks were real-life functions and activities. On the one hand and as a result of these tasks reflecting viable situations in life, the DA rubrics were task-driven to measure the abilities of speaking Arabic in realistic scenarios and not only to perform better in a test. On the other hand, the performance accuracy required for each task at each proficiency level was identified by recognizable descriptors. These descriptors were for (a) lexical control, (b) syntactical control, (c) sociolinguistic control, (d) delivery, and (e) text type (length of utterances). Therefore, the specific descriptors for these accuracy factors made the DA process construct-driven as well (Anton, 2009; Messick, 1994).

The Dynamic Assessment Rubrics Form was used with task-based language instruction in a classroom setting, and the teacher-researcher used them in two situations: (a) during the students' actual collaboration when they needed the teacher's assistance and (b) when the groups' reporters presented their products. During these opportune moments, the teacher-researcher could cause learning by providing the calibrated scaffolding. The observers used the same form while observing the teacher-researcher interacting with students in their ZPDs and group-ZPDs. Using The Dynamic Assessment Rubrics Form (Appendix D) helped the teacher-researcher to conduct reliably recordable tracking for students' progress indirectly and seamlessly and then compared his form to the observer's to measure the interrater reliability of using the Dynamic Assessment Rubrics Form.

The Validation Process of Rubrics

The teacher-researcher had used the same standards of the Dynamic Assessment Rubrics Form from 2004 to 2007 to conduct formative assessment for students attending the Arabic Basic Course. In addition to obtain feedback from assessment scholars, peers,

and students to validate these dynamic-assessment rubric, a form containing the deconstruction of only one accuracy factor, “structural control” (Child et al., 1993) as shown in Appendix D, was used with numerous students. This piloting of the rubric soon indicated that following the same steps in deconstructing the other accuracy factors was necessary for the teacher to have them ready for all possible emerging performances in classrooms or the initial interventionist DA of each student. These other factors, although not included in this study, are “lexical control,” “socio-cultural control,” “delivery,” and “text type” (Child et al., 1993).

Trying out the rubric form for these accuracy factors in the past showed that they would be closer to being task-driven during interventionist DA interviews (Anton, 2009; Bachman, 2002; Messick, 1994), because in such situations the teacher-researcher would evaluate the students’ performance more holistically. During instruction, however, the rubric form would become construct-driven as the teacher-researcher would be rather focused on evaluating the students’ performances against specific ILR descriptors, that is, the designing process of the classroom tasks would make the DA form task-driven; whereas recording the teacher-researcher’s mediation on the form would make the same rubrics construct-driven by evaluating the students’ performance against a specific ILR descriptor under one of the five accuracy factors mentioned above.

The teacher-researcher conducted a breakdown of the ILR criteria first in 2005 for the purpose of standardizing the formative speaking tests of the Arabic Basic Course while working as a Chairperson of an Arabic department. He trained all certified testers in the same school at that time on using the newly developed scale to raise its inter-rater reliability. The newly developed formative scale by deconstructing the ILR scale had

been used for few years, and proved to enjoy a high level of face validity and inter-rater reliability. Therefore, the teacher-researcher used the same tried-out descriptors for the purpose of devising the rubrics used in this study.

The remaining part of this section lists the forms, questions, and questionnaires that the teacher-researcher used during the three phases of this study. The forms used in this study: (a) Biographical Background Questionnaire (Appendix B), (b) Intellectual Styles Questionnaires (Thinking Style Inventory and Myers-Brigg Type Inventory), (c) the Dynamic Assessment Rubrics Form for Teachers (Appendix C), and (d) The Dynamic Assessment Rubrics Form for Observers (Appendix D). The descriptors for the speaking proficiency sublevels were deleted from Appendix C to save space on the form for the teacher-researcher who was intimately familiar with these descriptors. The teacher-researcher however used either version of the Dynamic Assessment Rubrics Form (Appendix C or Appendix D) occasionally to record his evaluation of the students' structural control only for the purpose of this study.

The version in Appendix D was the same as the other in Appendix C with the exception of including the deleted descriptors one time at the top row of the form so that the observers can refer to them when necessary. The teacher-researcher had intimate understanding of these descriptors and did not need to reread them while teaching in the classroom. The forms reflecting the dynamic assessment rubrics that were used during the instruction hour of the DA-phase, and then the teacher-researcher wrote entries in his teacher journal. The teacher-researcher interviewed each observer immediately after finishing the teaching of each lesson to discuss his or her entries on the form and to obtain his or her feedback on the DA/TBLI instruction.

Guiding Questions for Interviewing the Observers

The teacher-researcher interviewed observers immediately after each lesson. To find answers to the pertinent questions of this study, he used the following guiding questions.

1. What is your perception about the diagnostic abilities of the DA/TBLI instruction?
2. How practical is using the DA rubrics in class while teaching?
3. Do you think the DA process made a difference in students' learning/performance during your classroom observation?
4. Do you think teachers need training on using the Dynamic Assessment Rubrics Form before using it in classrooms?
5. Do you think teachers need training on the process of DA/TBLI instruction?
6. Is there any other information you would like to share with me about the use of DA/TBLI instruction?

Guiding Questions for Interviewing Students

The teacher researcher interviewed students immediately after the post-interventionist DA that he conducted for them individually. The following are the guiding questions for these interviews for the purpose of answering question 4 of this study.

1. What is your perception about the diagnostic abilities of the DA/TBLI instruction?
2. Do you think the DA/TBLI instruction made a difference in your learning/performance of Arabic speaking?

3. Did you benefit from the hinting process that was done with other students in your group?
4. Do you feel you had enough input on the subsequent lesson planning of the DA phase?
5. What do you think could be done to improve the process of DA/TBLI lessons?

Guiding Questions for the Teacher Journal

The teacher-researcher used the following questions to guide his entries in a teacher journal daily after each lesson. These questions were designed to verify the information collected from each observer's interview and to prompt the teacher-researcher to evaluate his agreement or disagreement with them.

1. How practical was the use of the DA rubrics?
2. Were the gradual hints used successfully?
3. Was the formation into small groups successful for the task?
4. Was the reading and listening material used suitable for the task and for the students' current proficiency level?
5. How can I use the collected diagnostic information in my subsequent lesson planning?
6. Which student showed progress in their structural control today in comparison to previous lessons?

The Ten 5-Point Scales

This survey was used to quantify the students' responses during the interview as a verification method. Both the interviews and the survey were designed to answer question 4 of this study.

1. The DA/TBLI instruction method is an effective classroom approach for language learning.
2. DA/TBLI instruction is capable of diagnosing each student's language needs on a daily basis.
3. The hinting process helped me overcome my personal language difficulties.
4. The hinting process that I experienced improved my speaking ability in Arabic quickly.
5. I would recommend DA/TBLI instruction for other language students.
6. Knowing the ILR standards helped me understand what I need to do to improve my speaking abilities.
7. Collaborating with other students to deliver a measurable product provided me with a great learning environment.
8. Following other students going through the hinting process helped me learning or overcoming my own personal difficulties.
9. Using DA/TBLI instruction in the classroom was practical and enjoyable.
10. Please use the space provided below to enter any additional information.

Use of Assessment

To identify the students' proficiency levels, this study included OPI and "Interventionist" DA sessions (Grigorenko & Sternberg, 2002; Poehner, 2005) at the

beginning and again at the end of this case study for each participant. The interventionist DA followed the same structure and tasks of the Oral Proficiency Interview, but it provided the students with the dynamic-assessment scaffolding too. The Interventionist dynamic-assessment sessions were for the purpose of diagnosing each student's strengths and weaknesses so that the DA phase was tailored to his or her needs for accomplishing proficiency level 2 at the end of Semester III. In this study, the DA Interventionist approach was administered for each student in the same structure of the institute's Oral Proficiency Interview as explained later in the next section of this chapter. The Oral Proficiency Interview was the "static" psychometric test, and it was only capable of measuring mature abilities. This study compared the Oral Proficiency Interview to the process of dynamic assessment to learn how they differed in evaluating Arabic.

Replicating the approach of a previous study, Poehner (2005), the teacher-researcher conducted the "interventionist" DA sessions at the beginning and the end of this study. The interventionist DA was conducted at the end to check their parallel reliability against the students' concurrent Oral-Proficiency-Interview tests. Certified Oral Proficiency Interview testers conducted the Oral-Proficiency-Interview tests prior to the instruction segment of this study. Comparing the results of the Oral Proficiency Interview and the dynamic-assessment interviews at the beginning and the end of this study contributed to answering this study's questions. The daily DA evaluations assessed continually the present proficiency levels ("i") of every student and group of students so that suitable input material at ("i+1") would be selected for the daily activities.

This research used peer observation by certified-Oral-Proficiency-Interview testers, interviews, and DA evaluation surveys to record the students' and the testers'

reaction to the dynamic-assessment approach. Certified-Oral-Proficiency-Interview testers conducted the peer observation to fill out the Dynamic Assessment Rubrics Form; comparing their Dynamic Assessment Rubrics Form to the one used by the teacher-researcher for the same lesson measured the reliability of the dynamic-assessment process in the classroom. For example, the observer used the same Dynamic Assessment Rubrics Form to record the immature abilities based on the teacher-researcher's gradual hints provided to students during the observed lesson. Comparing the teacher-researcher's entries on the form with those entered by the observer reflected on the fidelity of using these rubrics.

The reiterative cycle of the peer observation mentioned above assessed the interrater reliability between the teacher-researcher and all the observers involved. The process of using ILR certified-Oral-Proficiency-Interview testers, the teacher-researcher included, enhanced the validity and the reliability of using the Dynamic Assessment Rubrics Form and consequently the whole classroom dynamic-assessment process (Bachman, 1990). To answer the study's question about the observers' experiences and perception of the DA/TBLI instruction, the teacher-researcher interviewed them to elicit their feedback on the validity, reliability, and practicality of the DA/TBLI instruction and the Dynamic Assessment Rubrics Form including their suggestions for improving it.

The Oral Proficiency Interview

The Oral Proficiency Interview (OPI) is the instrument used by the U.S. Government to evaluate the speaking functional abilities with a foreign language in real-life situations. Both DLIFLC and the Foreign Service Institute used the term "proficiency" to reflect the ability of a target-language user to function with the language

in a real-life situation. The definition overlapped with the term “authentic assessment” (Baron & Boschee, 1995; Foster & Skehan, 1996) to the greatest extent, because the learners’ speaking abilities during the OPI were performed during their communication of a real-life function. These functions included but were not limited to narrate in all time frames, providing physical description, reporting facts, defending a personal opinion, hypothesizing and participating in role-play for survival situation or unfamiliar situation in the target culture. The term used in the Oral Proficiency Interview Manual (Child et al., 1993; DLIFLC, 2010) for each one of these real-life functions was “task.” Although this term did not meet the criteria of a “task” as used in task-based language instruction and as mentioned in Chapter II, these functions called tasks in the Oral Proficiency Interview were real-life products as required in the definition of authentic assessment (Baron & Boschee, 1995). The teacher-researcher integrated both definitions in designing the task-based activities for this study, that is, tasks crafted for this study were closer to Bachman’s (2002) definition for task-based language instruction performance assessment.

There were different real-life functions for the examinee to perform at every proficiency level, which qualified the Oral Proficiency Interview as a task-based language performance assessment (Bachman, 2002). The Interagency Language Roundtable scale described the accuracy level for performing each one of these tasks for every proficiency level. The descriptors of the different proficiency levels were documented in the Interagency Language Roundtable scale. These descriptors were the set standards that reflected the abilities of performing real-life tasks in all modes of listening, reading, speaking, and writing. That is, the ILR scale contained the rubrics of

the task-based-language performance-assessment functions for each proficiency level (Child et al., 1993).

The Oral Proficiency Interview used the descriptors listed in the ILR scale as rubrics for evaluating the examinees' speaking proficiency levels, and these proficiency levels were categorized hierarchically by the same labels for the other target language modes: listening, reading, and writing. These different proficiency levels ranged from no functional ability in the target language to that of a well-educated native speaker. The proficiency levels were coded and labeled as follow for Speaking since 1985 (ILR, 2013a): "S 0," "S 0+," "S 1 ," "S 1+," "S 2," "S 2+," "S 3," "S 3+," "S 4," "S 4+," and "S 5." Proficiency level S 0 meant no functional ability in the target language, whereas S 5 reflects a performance equivalent to the abilities of well-educated and articulate native speakers (DLIFLC, 2010; ILR, 2013a).

Practicality referred to cost-effectiveness or to the ease of scoring and administering a test (Child et al., 1993). The requirements of conducting OPIs were very simple. The institute provided quite rooms with a recording device. Interviews were recorded digitally, and the rating of the two needed testers for every interview was entered on a simple rating form. Only very few other documents were used: the ILR Skill Level Descriptions, a set of role-play cards, a set of tester cards, in addition to the rating sheet for each tester. On the back of this rating sheet form is the Rating Factor Grid for testers to review before finalizing their decision for awarding the rating. Testers needed from 15 to 20 minutes to fill out this form (DLIFLC, 2010).

The Oral Proficiency Interview would take on average from 15 to 45 minutes depending on the interviewee's proficiency level and the testers' concurrence of the

hypothetical working level. Sometimes, in difficult assessment situations, the interview would last longer than the scheduled time due to either one of the two reasons mentioned above. The two testers would spend about 10 minutes to finalize their nonconference rating of the interviewee's performance by reviewing the ILR standards provided on the back of the rating form. Then, each tester would fill out the rating sheet to enter their final rating, and they were encouraged to write a justification for their rating in the space provided. They were required to sign and date this form (DLIFLC, 2010).

The Defense Language Institute Foreign Language Center conducted about 7,000 OPIs a year (DLIFLC, 2013b), which reflected that this test had matured to a very practical test since 1970. The American Council on Teaching Foreign Language depended on the Defense Language Institute Foreign Language Center's testers to help them in conducting their quality control process by "third rating" the interviews conducted by their certified testers. "Third rating" was a term used in DLIFLC for the quality control conducted by a third tester who would listen critically to the recording of a test to make sure that the rating was accurate and to make sure that testers would comply with all standards. Third raters used a special form to record their findings, and this form was called the Third Rater Analysis Form (DLIFLC, 2010).

These testers at the Defense Language Institute Foreign Language Center were recertified annually by going through evaluative training in addition to having a one-on-one training called a tester support session. These tester support sessions were usually conducted by getting the tester to third-rate one of his old tests. The two testers who conducted the recorded test listen to it critically with their trainer to reflect on their performance. Having testers whose interrater reliability was heightened by such a strict

quality control process and tester training sessions was a great asset to this study. Their intimate understanding of the ILR descriptors would be transferred easily to the ILR-based Dynamic Assessment Rubrics Form. They would need only a short training to understand the new sublevels of this form and how to enter the number of hints used by the teacher-researcher on it while observing his lessons.

OPI Structure

The Oral Proficiency Interview was a conversationalist interview that consisted of the following stages and was administered by two testers: (a) warm-up stage, (b) reiterative stage, and (c) wind-down stage. The rating focuses only on the reiterative stage, which included “level checks” and “probes.” The warm-up stage helped the testers to collect information about the examinee to use during the Reiterative part, and it was used to help the interviewee relax before he or she started the ratable segment of the Oral Proficiency Interview. During the warm-up part, testers hypothesized the working level during the reiterative stage, which was the level each tester intended to award by the end of the interview. Basically, testers used the reiterative stage, the ratable part of the interview, to prove to the system that their hypothesized level was correct (DLIFLC, 2010).

The reiterative stage consisted of “level checks” and “probes.” The level checks were the tasks required for the working proficiency level (hypothetical level), and the probes were tasks from the higher proficiency level. On the one hand, the purpose of eliciting level checks was for the interviewees to demonstrate through their performance of the tasks their abilities of meeting the standards described for the hypothesized working level, that is, level checks established the “floor” (DLIFLC, 2010) for the

examinee's proficiency level; the floor was the speaker's proficiency level. The examinee was rated at the end of the interview by the proficiency level of the floor. On the other hand, the purpose of eliciting probes was to make sure that the examinee cannot meet the standards required for the higher proficiency level. This meant that eliciting the probes established the candidate's "ceiling" (DLIFLC, 2010). For an Oral Proficiency Interview to be ratable, testers had to elicit level checks for all the tasks necessary for the examinee's proficiency level and to elicit two probes from the higher level to confirm that the examinee cannot satisfy its standards (DLIFLC, 2010).

If the examinee fulfilled all the descriptors of the checked level, the floor, then testers rated his or her speaking abilities with the floor's proficiency level as coded on the ILR scale. The plus levels had been included in the ILR since 1985 (ILR, 2013b) to reflect the substantial departure of the examinee's abilities away from one base level toward the higher proficiency level. To award the "plus" level of any base level, testers elicited probes four times and not only twice as described above for the base levels. The reason was to give the examinee further opportunity to perform at the higher level to show his or her inconsistent performance at it or his or her substantial improvement over the standards described for the lower base level (DLIFLC, 2010).

Tasks and Their Coverage

Each proficiency level had its own tasks that a speaker needed to perform successfully with the accuracy level described in the ILR guidelines to be rated by it (DLIFLC, 2010, 2013b). The tasks of proficiency Level 1 of speaking included the eliciting of simple short conversation about a daily survival need, a role-play for a survival situation, and the examinee's ability of asking simple questions. The tasks of

proficiency Level 2 of speaking included narrating in all time frames, describing a physical object, giving instructions or directions, a role-play about a survival situation with complication, and reporting facts on current events (DLIFLC, 2010; ILR, 2013a). Speakers at Level 2 (L2) were able to speak with minimum cohesive utterances, and their longer utterances were coherent (ILR, 2012a). The learner would speak with confidence but not with facility at paragraph-long utterances (ILR, 2012a). Although the speakers' mistakes would be frequent, their basic grammatical structures would be controlled typically. Unlike the speakers at Level 1, the speakers' delivery at Level 2 would be understood to all natives including those who were not used to dealing with foreigners (DLIFLC, 2010; ILR, 2012a). Their lexicon included concrete vocabulary items, and this was one of the limits that separated them from Level-3 speakers (DLIFLC, 2010; ILR, 2013a) who also used abstract and specific words.

The extended discourse utterances at Level 3 were cohesive in performing the following tasks: giving an opinion on societal issues, discussing or commenting on an abstract topic, and a role-play for an unfamiliar situation in the target culture (DLIFLC, 2010; ILR, 2013a). Although these three tasks would be elicited at levels from L3 to L5, the scope of abilities would escalate from societal at L3 to philosophical at L4 and L5. Performance would improve in all the five accuracy factors used in the ILR guidelines for all levels: (a) lexical control, (b) structural control, (c) sociocultural control, (d) delivery, and (e) text type (length of utterance). The abilities of every higher proficiency level subsumed all the lower levels (DLIFLC, 2010; ILR, 2013a).

Reliability

OPI had high face validity (Bachman, 1990, 2002; Bachman & Palmer, 1996; DLIFLC, 2010). Not only DLIFLC alone conducted more than 7,000 interviews per year, the ILR as the rubrics used for evaluating the interviewees' performances had proven to be consistent in discriminating between performances. The Government started using the ILR guidelines during the war with Japan after realizing the dire need for a reliable and valid instrument to evaluate foreign language abilities (ILR, 2013a). First, the ILR started by evaluating foreign languages without specifying the different modes of listening, reading, speaking, and writing (ILR, 2013a).

Then, modifications were made to the ILR to separate the standards for these four modes in 1968 (ILR, 2013a). As mentioned above, the fine-tuning of the ILR guidelines continued until its modification in 1985 to include the "plus" levels to reflect the substantial improvement of the speaker's abilities over the descriptors of the base levels in the ILR. These rubrics of the task-driven OPI (Messick, 1994) had been evaluated holistically (Bachman, 1990) for decades, and therefore OPI has both inter- and intrarater reliability as the result of the moderation that was maintained through several measures.

In a study (Bienkowski, 2013) conducted on 709 students of Modern Standard Arabic, French, and Spanish at proficiency levels from Level 0+ to Level 2, Bienkowski (2013) tried to answer the following questions (a) Are the ILR Can Do Statements measuring perceived language proficiency consistently and accurately for all Special Operations Forces Teletraining System students? (b) Are the Can Do Statements related to similar constructs such as students' confidence in their ability to perform language tasks? The data collected were analyzed according to the classical test theory, item

response theory, and correlation with other perceived theories. In general, findings suggested that the Can Do Statements subscales were measuring consistently the same construct. The internal consistency reliability (Cronbach's alpha) for the different proficiency levels were: .88 for Level 1, .90 for Level 2, .87 for Level 3, and .82 for Level 4. For the second questions, the study found a strong correlation ($r=.79$) between the Can Do statements and the assigned course level and the perceived proficiency level from the pretraining survey. In the coming section of this chapter, the procedures followed to allow individuals to acquire shared understanding of the performance standards are reviewed. The purpose of this shared understanding was elevating the inter-rater reliability and the validity of the Oral Proficiency Interview.

Training Raters

After their selection as prospective Oral-Proficiency-Interview testers, all raters went through 3-week certification training for 8-hour days (DLIFLC, 2010). During this very intense training, raters went through about 25 Oral-Proficiency-Interview ratings whether directly or indirectly. The word directly meant that the rater would be one of the two testers conducting an informal Oral Proficiency Interview on a volunteer from outside the workshop that could be at any proficiency level. Then, all participants rate the examinee blindly in a nonconference manner to discuss their ratings thereafter. At the end of this workshop, participants who would enjoy consistent successful rating were certified provisionally. Their certification would become complete after spending a probationary period during which their performance would be monitored closely.

In addition, testers would receive recertification training annually for 2 full days to make sure that they maintained their common understanding of the rubrics. To make

sure that this understanding was maintained, testers would be called for up to five times a year for a process called “tester support.” In this process, the two testers of any particular Oral-Proficiency-Interview examination would be called in to listen critically to the tape of their own test followed by a discussion about their elicitation techniques, their ratability of the test, and their rating accuracy. In addition and for the purpose of raising the interrater reliability, all OPIs that would end up in a split rating receive a blind third rating. A split rating meant that the two testers did not agree on their nonconference rating, and the blind third rating means that the third rater listens critically to the recording without knowing the ratings of the initial two testers (DLIFLC, 2010). Two testers have to agree on a particular rating to hold. The two testers who produced the split rating are usually called in for tester support. Statistics are done on all tests every year to make sure that the coefficient for the inter-rater reliability is acceptably high.

Validity

This section contains the internal characteristics of the OPI. One of the pluses of the OPI, which would answer the concern about the validity of Performance-Based Assessment (Bachman, 2002), was the Oral Proficiency Interview’s interwoven structure as being both task-driven and construct-driven simultaneously (Messick, 1994). These two domains were included by prescribing certain tasks as required at each proficiency level, and the performance accuracy for each of which was described as well. For example, the following were three of the L2 tasks: narration in the present, narration in the past, and narration in the future. These were both tasks and constructs at the same time, because the ILR rubrics explained in more details the expected accuracy by which the performance of these real-life products, functions, or tasks should be. Moreover, the

issue of “representativeness” raised by Bachman (2002) was solved in the OPI process and using the ILR scale. By representativeness, he questioned if the selected task in the task-based language performance assessment would represent all the tasks that a person could possibly do in daily real-life situations. Representativeness was solved in the OPI by complementing the definition of a task with the dimension of functioning (DLIFLC, 2010). The Interagency Language Roundtable scale describes the accuracy by which the OPI tasks or functions should be performed.

Representativeness would have been a problem, if the task were to narrate in the past about a certain event, but the ILR tasks measured people’s ability to function with the target language in a real-life task, that is, the testers would prompt the examinee to narrate about any random event. Evidently, one event would never represent all possible developments in the past that people would need to tell someone about. Rather, the task in the Oral Proficiency Interview would be to function with the target language to narrate in the past about any event that happened to the speaker or to someone else. A well-trained tester would make certain that the flow of the interaction during the Oral Proficiency Interview would lead to prompting the examinee to narrate about any of his or her past events to elicit unrehearsed performance. Avoiding rehearsed material would help in evaluating the candidate’s real abilities in the target language. Reviewing all the tasks at all the different proficiency levels clearly would show that the same principle of being task-driven and construct driven at the same time apply to all of them. Further, a component of performance-based assessment and authentic assessment would be problem solving a real-life situation (Johnson et al., 2009).

The fact that the learner would be carrying out a conversation with two natives who were interacting about unprepared and unrehearsed topics was a problem-solving situation in itself, which consequently would qualify the Oral Proficiency Interview as an authentic assessment test as well. Besides, the tasks for all proficiency levels would include a role-play that would introduce a problem for the learner to solve as it would happen in reality (Child et al., 1993; Foster & Skehan, 1996). A role-play would start at L2 by introducing a little complication to a survival situation. At L3, the interviewee would need to function as described in the ILR scale to solve an unfamiliar situation in the target country to reflect his or her cultural awareness and communicative competence (Canale & Swain, 1980) as needed at this level. The same would apply to the role-plays at both L4 and L5; examinees would need to perform on two role-play situations. One role-play would be for a formal situation and the other would be for an informal situation. An example for the formal situation would be prompting the examinee to address either one of the two houses of Congress. In an informal situation, the examinee would need to function appropriately with the TL in the target culture to advise a very close friend or a relative who would be facing a life crisis.

Sensitivity to Instruction

The Oral Proficiency Interview was a proficiency-based evaluation of the speaking ability, and by definition it did not measure the mastery of a certain curriculum. Rather, it measured the examinee's functional ability with the target language in real-life (DLIFLC, 2010) situations. Therefore, instruction in the classroom at the Defense Language Institute Foreign Language Center aimed at raising students' proficiency through experiential and student-centered approaches such as task-based language

instruction. Task-based language instruction (Ableeva & Lantolf, 2011) would promote student-centeredness by collaborating in small groups to process multimodal input to prompt students to use their critical thinking to generate a product for a real-life situation (Ableeva & Lantolf, 2011). Although both the Oral Proficiency Interview and classroom instruction at the Defense Language Institute Foreign Language Center would address the learners' abilities to function with the target language in daily-life scenarios, students could never know what to expect to discuss during the test, which would raise the level of Bachman's (2002) representativeness for the Oral proficiency Interview.

Consequential Validity

Graduates of the Defense Language Institute Foreign Language Center who accomplished the school's objective of Level 2 in speaking at the end of the Basic Course would receive a bonus pay in addition to their regular salary for as long as they would maintain their proficiency level. Moreover, their proficiency level as determined by the Interagency Language Roundtable scale would be detrimental in the jobs that they would do in the different military services and consequently chances for possible promotions and retention pays later. In some services at times, the consequences were dire for students who did not succeed to meet the language training objectives. Students would be motivated consequently to excel during their course of instruction and to graduate with the best results possible. This instrumental motivation would drive them to endure the increasing challenges and demands of the Arabic Basic Course.

Fairness and Equity

To assure that all students would receive a rating that would reflect their real abilities without any confounding effects, teachers could not test their own students.

Teachers in management position could not test students. The pretest instructions would inform examinees that the process would not rate their ideas or attitudes on any issue; rather, their use of the target language to express their opinion would be only the matter. To make sure that students' performances were not psychologically impacted, the same instructions would inform them before the beginning of the interview that the testers could change any sensitive topic for the examinee (DLIFLC, 2010). Before the beginning of the Oral Proficiency Interview, students would be placed in a controlled waiting area where they could not mingle with others who had just finished their Oral Proficiency Interview.

In general, the ILR included 11 proficiency levels for each language skill (listening, reading, speaking, and writing). The Defense Language Institute Foreign Language Center's graduation standards were Level 2 in listening, Level 2 in reading, and Level 1+ in speaking. Writing was used in the program as an enabling skill only, and, therefore, writing abilities were not evaluated at the end of the Arabic Basic Course. As the Oral Proficiency Interview was used to evaluate Speaking, the Defense Language Proficiency Test V (DLPT V) was developed to evaluate the proficiency of listening and reading according to the Interagency Language Roundtable scale. The Interagency Language Roundtable standards described the abilities of listeners and readers at each proficiency level, and, therefore, the developers of the DLPT V needed to include passages whose difficulty level would be congruent with the ILR criteria for listeners and readers at the measured proficiency levels. This congruency meant that the difficulty level of the passages used are suitable for the readers and listeners as described for a particular proficiency level on the ILR scale. The language features included in these

passages could be processed by readers and listeners whose abilities match the descriptors of the ILR guidelines for each proficiency level measured by the test.

Text Typology

Considering that the Interagency Language Roundtable scale described the abilities and inabilities of listeners and readers and not the difficulty levels of passages, the developers of the Defense Language Aptitude Test V (DLPT V) referred to the Child's (1987) work on "text typology." Text typology (Child, 1987; Child et al., 1993) as known in the U.S. Government defined a text as a string of connected or disconnected words that were spoken or written, and it described the factors that made one text more difficult or easier to process for a listener or a reader. These factors were (a) the topic, (b) text type (editorial, advertisement, announcement, etc.), (c) text mode (the author's or the speaker's intent), (c) schemata (linguistic, cultural), (c) vocabulary (concrete, abstract), (d) syntax, (e) register, and (f) style. The pivotal factor of these eight factors was the text-mode element. Text mode in this context meant the author's or the speaker's intent, and these intents actually were used in labeling the texts' different difficulty levels. From easier to more difficult to process, the names of these modes were (Child, 1987, 1998, 2001) (a) "Enumerative," (b) "Orientational," (c) "Instructive," (d) "Evaluative," (e) "Projective," and (f) "Stylistic." Each one of these modes had a full description for the other factors in a document called Density and Syntax (Child & Lowe, 1998).

The difficulty level of each text mode matched the abilities described for listeners and readers in the ILR proficiency levels. For example, a listener or a reader at levels L0+ or R0+ would be able to process with acceptable accuracy and comprehension a passage in the enumerative text mode as described in the Density and Syntax Chart (Child &

Lowe, 1998). The same would apply for each proficiency level in the order listed above until the highest stylistic text mode would match the abilities of Level 5 in both modes of listening and reading. The plus-levels as described in the ILR scale would match texts that fell in between two modes, these were called “mixed modes” (Lowe, 2000). The compatibility of text typology with the Interagency Language Roundtable empowered foreign language educators in different areas of the field. These areas were curriculum development, test development, diagnostic assessment, and passage selection for the classroom supplementary material. The reason was that the ILR had established high validity and reliability and the development of text typology was based on the ILR descriptors and by the same developers (Child, 1987, 1998, 2001)

The Defense Language Institute Foreign Language Center had used the parity between text typology and the ILR guidelines in many aspects of its language teaching mission. One of these aspects was the development of tests such as the Defense Language Proficiency Test V. A second aspect of using text typology was curriculum development or in preparing classroom supplementary material. The curriculum developer would need to hypothesize the students level at each point in time for each lesson in the textbook so that he or she would follow the formula of “i+1” to select a suitable text (Child & Lowe, 1998). The teacher-researcher of this study needed to follow the same principle to select a comprehensible passage for the participants’ daily lessons that included the language features he intended to teach. Therefore, the teacher-researcher used dynamic assessment in this study to evaluate the students’ current proficiency levels almost daily, and then based on their determined proficiency levels, he selected the suitable difficulty level for the text used in the subsequent lessons.

The Teacher-Researcher

The teacher-researcher has been working with the Defense Language Institute Foreign Language Center since 1991. He worked in the following assignments during his tenure in the institute: Arabic Instructor, Arabic Video Tele-training Instructor, Arabic Team Leader, Diagnostic Assessment Specialist, Diagnostic Assessment Branch Chief, Arabic Curriculum Writer, Tester Trainer, Arabic Department Chairperson, Dean of Educational Support Services, Faculty Development Specialist, and Information Technology Officer. In addition to these positions, the researcher had worked in DLIFLC as a teacher since 1991, as a certified-OPI-tester since 1997, and as an Arabic Master Tester since 2001. As a tester, tester trainer, master tester, and a Diagnostic Assessment specialist, the teacher-researcher had worked very closely with interpreting the ILR-level-descriptions for Speaking, conducting OPI interviews, and executing third ratings. The third-rating of an OPI was a blind rating for a test by a third certified tester, and it was conducted mainly for quality control purposes.

The countless number of OPIs conducted by the teacher-researcher in the last 16 years had given him priceless opportunities to experience numerous profiles of Arabic speakers at all proficiency levels as measured by the ILR scale. In addition to these interviews, the teacher-researcher had the experience of conducting OPI-like interviews as one of the main segments of the Defense Language Institute Foreign Language Center's diagnostic-assessment-three-skill interviews. During these speaking segments, the teacher-researcher had tried to identify the Arabic features needed for the interviewees to perform at the next ILR-proficiency level. Based on the finding of the three-skill interview, the next step was to develop a learning plan for the interviewee for

the purpose of helping her or him advance to the desired level on the ILR scale.

Providing a learning plan for the interviewee included a plan to advance in listening and reading on the same scale as well. To identify the Arabic features needed in these two skills, the teacher-researcher used text typology to rate and select passages for both modes; these passages were used also during the three-skill interview to identify the interviewee's abilities and inabilities in both listening and reading according to the ILR scale.

The teacher-researcher's extensive experiences of rating and selecting Arabic passages enabled him to select the input material used in designing tasks for the daily lessons of this research's DA phase. The teacher-researcher was able to select interesting passages for the targeted tasks on the ILR scale by allowing students to share the planning process, and these passages were at a difficulty level suitable for their present proficiency level as evaluated through the dynamic-assessment process and the Pre-DA data. The teacher-researcher's sensitivity to the students' utterances as an OPI tester empowered him to evaluate their present proficiency level and filling out the Dynamic Assessment Rubrics form effortlessly and seamlessly. These experiences enabled him to collect and analyze the data reliably to answer the questions of the study.

Research Questions

The following is a restatement of the study's questions followed by an explanation of how the teacher-researcher analyzed the collected data to answer them.

1. What is the change in the structural control of Arabic speaking based on DA/TBLI instruction?

2. How do OPI without DA assistance and OPI with DA assistance compare relative to the evaluation of Arabic speaking?
3. How do the experiences and perceptions of DA/TBLI instruction compare between teacher-researcher and OPI testers?
4. What are the student perceptions of the DA process?

Data Analysis

To answer question 1: The researcher compared results of pre- and post-OPIs and pre- and post-interventionist DA interviews relative to the structural control components of Arabic speaking. More specifically, the researcher, first, compared the pre-OPI with the post-OPI to investigate if there was any recordable improvement on this psychometric static test. These results only reported the proficiency level of each examinee as coded on the Interagency Language Roundtable scale. By comparing the pre-interventionist DA with the post-interventionist DA, the researcher was able to identify changes in the structural control components of Arabic speaking. The details of these possible changes were tracked by the other kind of DA, the interactionist DA, to examine the language-acquisition developmental progress that had led to the changes in the post-interventionist DA. Finally, comparing the pre-OPI to the pre-interventionist DA and the post-OPI to the post-interventionist DA helped the researcher to examine their congruency in rating the examinee's proficiency level, that is, this process examined their parallel reliability.

To answer question 2: The researcher compared OPI without DA assistance to OPI with DA assistance relative to the amount of diagnostic information provided regarding Arabic-speaking ability. Comparing the OPI results to the interventionist and interactionist DA results on its rubrics shed light on their limitations and abilities to

provide the learner with diagnostic feedback accurately. Comparing them showed the potential of both assessment types to promote learning and language acquisition. The teacher-researcher interviewed students in the post-DA phase to know how effective the DA process was in diagnosing their needs for the planning of subsequent lesson planning. Interviewing students contributed their perception of the effectiveness of DA/TBLI instruction, which was the interactionist DA, in improving their Arabic speaking ability.

To answer question 3: The researcher determined the nature and perception of DA/TBLI between the teacher-researcher and the Oral-Proficiency-Interview (OPI) testers by reviewing the teacher journal and the observers' interviews. Interview responses from OPI testers, the observers, were recorded and transcribed. Both interview transcripts and the teacher journal reflections were coded and reviewed for emerging themes. Themes then were analyzed in relation to teacher-researcher and OPI testers' experiences and perceptions of DA/TBLI.

To answer question 4: the teacher-researcher interviewed students during the post-DA phase. These interviews were recorded, and their transcripts were reviewed for emerging themes. Themes were analyzed then in relation to student perceptions of the DA process. More specifically, the researcher got their perceptions on the diagnostic abilities of the DA/TBLI process, the hinting process, the sufficiency of their input to the subsequent lesson planning, and the effect of DA on their Arabic speaking abilities. In addition, students' responses to the ten 5-point scales were evaluated to quantify their perception for the DA/TBLI instruction including its different techniques.

CHAPTER IV RESULTS

Overview

This chapter contains the findings of this study, and is divided into seven parts. The first two parts are for the participants and the design overview. The section for the participants summarizes their background, selection, and their aliases in this study. The section on the design overview summarizes the design of this study and the method used in collecting and analyzing the data. The following four parts report the results for the study's four research questions. The last part is a summary of all the results mentioned in this chapter.

Design Overview

The study was conducted in three phases: pre-DA, the DA, and post-DA. In phase one, students answered questionnaires of Myer-Briggs Type Indicator (MBTI), Thinking Styles, and their biographical background. The teacher-researcher conducted a pre-interventionist DA for these selectees before they started the next phase. In phase 2, students received a lesson of one hour a day during which they were exposed to the DA/TBLI process. The teacher-researcher and his observers used rubrics designed especially for this approach; these rubrics were included in a form called the Dynamic Assessment Rubrics Form. The teacher-researcher interviewed each observer after each lesson to inquire about their experiences and perception for the DA/TBLI approach. He entered his own experiences and perception into a teacher journal after each lesson during the DA phase.

In phase 3, two different certified testers conducted the post-OPIs for each participant, and the teacher-researcher conducted the post-interventionist DA for each

student. He also interviewed each student after the post-interventionist interviews to inquire about their perception of the DA/TBLI approach. Additionally, each student responded anonymously to a survey of ten 5-point scales online or on a hard copy.

The following answers for this study's questions were determined by comparing students' evaluations during the pre-DA phase to those done during the post-DA phase, evaluating their progress during the DA phase, comparing data from all interviews with the teacher-researcher's journal, and by the results of the students' survey. The progress that occurred by using both techniques of the dynamic assessment was determined by comparing the number of hints used for the same language feature during both the pre- and post-interventionist interviews and during the interactionist-DA used during the DA phase. These hints graduated from being most implicit to being most explicit as follow: (a) not accepting the answer, (b) repeating the erroneous part, (c) repeating the specific erroneous utterance, (d) naming the grammatical feature, and (e) providing the student with the correct answer and its explanation.

Research Question 1

Research question 1 asked: What is the change in the structural control of Arabic speaking based on DA/TBLI instruction? Comparing the static and dynamic assessments conducted during the pre-DA phase with their corresponding interviews of the post-DA phase demonstrated improvement not only in the participants' structural control but also in their proficiency level on the ILR scale. Pre-OPI and post-OPI results also showed improvement for all students. Table 1 showed the results for the pre-OPIs, post-OPIs, and the official OPI conducted formally by DLIFLC as the exit static test of the Arabic Basic Course. Students received their official OPIs five weeks of instruction after this study.

Performance in these interviews was evaluated by the ILR scale as mentioned in Chapter III.

The descriptors for the structural control of Level 1+ on the ILR scale reflected accuracy in basic grammatical relations that was evident but not consistent. As explained further in the head of the first column of the Dynamic Assessment Rubrics Form (see Appendix D), the speaker at Level 1+ might exhibit the more common forms of verb tenses, for example, but might make frequent errors in formation and selection. This individual cannot sustain coherent structures in long utterances or unfamiliar situations. The speaker's references to person, space, and time were often used incorrectly. Improvement in these references to form and producing coherent long utterances were needed to advance to Level 2.

The descriptors of Level-2 speakers reflected noticeable advancement in the structural control of Arabic. A Level-2 speaker on the ILR scale and as used also for the third column of the Dynamic Assessment Form showed control of all tenses. His utterances were minimally cohesive. Also, the speaker's basic grammatical structures were typically controlled. His reference to person, space, and time were often used correctly. The speaker at this level could sustain coherent structures in longer utterances. The sublevel in between Levels 1+ and 2 was described in the second column of the Dynamic assessment Rubrics Form.

Before delving further into the relevancy between the ILR scale and the Dynamic Assessment Rubrics Form, it was important to note that though the ILR's proficiency levels were hierarchical, they were not of equidistance away from each other. The speaker had to improve exponentially in the six accuracy factors to advance from one

level to the next. Although all students started this study at Level 1+, as will be explained in more detail later, the interventionist interviews detected that they were at different proficiency sublevels in the range between Level 1+ to Level 2. That is, they had different profiles and some of them were more competent than the others as will be discussed in Tables 2 to 7 later.

The speaker who met the descriptors of column 2 of the Dynamic Assessment Rubrics Form showed control of all tenses most of the time. His or her utterances were minimally cohesive most of the time. His or her basic grammatical structures were typically controlled most of the time. The speaker's references to person, space, and time were often used correctly. The speaker at this sublevel could sustain coherent structure in longer utterances most of the time. Comparing this sublevel to Level 2 of the ILR scale showed that the only difference was the inconsistency of meeting the standards at level 2 as expressed by the phrase "most of the time" after each descriptor. The advancement from Level 1+ to 2 was mainly reflected in the speakers' ability to produce coherent long utterances with minimum cohesiveness. This minimum cohesiveness was made possible by referring often to person, space, and time correctly while having a sufficient control of the basic grammatical structures.

Table 1
Comparing the Results of the Pre-OPI with the Post-OPI

Students Name	Pre-OPI	Post-OPI	Official OPI
Jamal	Low 1+	H 1+	1+
Basem	1+	1+	1+
Ramzy	Low 1+	1+	2
Ibrahim	Low 1+	2	2
Hazem	High 1+	2	2
Salwa	Low 1+	1+	1+

Note. Ratings that improved in the post-OPI or the official OPI are in boldface.

According to data in Table 1, five out of the six participants, 83% of the students, showed improvement from pre- to post-OPI results. The OPI result for one of these students, Ramzy, improved even further in the official OPI later. OPI testers gave diagnostic feedback after informal interviews by mentioning if the rating was “high” or “low.” When they did not mention either one of these two expressions, they meant implicitly that the examinee demonstrated average performance of the awarded rating as it was described by the ILR scale. The problem with the two expressions of “high” and “low” was that they were not defined in any written document; they developed rather spontaneously among testers to approximate the students’ places in the ranges between any two levels in the ILR scale.

Tables 2 to 8 demonstrate the change in instructional control by comparing the pre- and post-interventionist interviews. Each one of Tables 2 to 7 was designed to show the changes between these two interviews for each student’s assisted features. The mean for each grammatical feature assisted for all students combined between the pre- and the post-interventionist interviews are compared in Table 8. The students’ changes observed demonstrated improvement in the structural control of all the features assisted. Before displaying and discussing the tables, it was important to note that students needed to perform the assisted features as described for the higher sublevel to advance to the second or the third column (Appendix D). Only two students started the DA phase fulfilling the descriptors of the second column, and the four remaining students started with abilities described for its first column. All students advanced to a higher sublevel and improved on their structural control of Arabic as shown in Tables 2 to 8 later.

The first column (see Appendix D) reflected the performance of Level 1+, whereas the second column's descriptors reflected the performance of a speaker almost at the threshold of Level 2. The third column described the performance of Level 2 on the ILR scale. The two students who started the DA phase under the second column (please refer to Appendix D) were Hazem and Ibrahim. The pre-OPI diagnosed Ibrahim as a "low" 1+ while the pre-interventionist interview and the daily interactionist DA during the DA phase evaluated Ibrahim as fulfilling the descriptors of the second column, that is, the results from the interventionist interview did not agree with the pre-OPI diagnostic rating. The tracking of students' progress during the DA phase will be reviewed individually in the next section by displaying and discussing one of Tables 2 to 7.

Basem

Table 2

Comparison of the Pre- and Post-Interventionist Hints: Basem

Language Feature Assisted	Hints		
	Pre-interventionist	Post-interventionist	Change
Adjectival Phrase	3	0	-3
Conjugating Past Tense	3	0	-3
Verbal Noun	2	0	-2
Long Utterances	Avoided	0	-5
Negating Nominal Sentences	5	0	-5
Using Present T After أن	Avoided	0	-5
Passive Voice	Avoided	0	-5

Note. The word "avoided" in all tables reflected that the student stayed away from the feature by overproducing other features. In this table, the word "avoided" was quantified as five hints. The "0" refers to independent performance.

Basem's performance and interactions with the teacher-researcher during the DA phase started by meeting the descriptors of the first column of the Dynamic Assessment Rubrics Form (Appendix D). As explained previously in this section, meeting the standards of column 1 indicated that he was just meeting the standards of Level 1+ on the Interagency Language Roundtable (ILR) scale. He started by avoiding the necessary long utterances needed for him to advance to the sublevel described in the second column of Appendix D. During these long utterances, he needed to control his speech using all the different tenses. He needed assistance, however, during the first 5 days of the DA phase with the basic grammatical features that were necessary for producing minimally cohesive sentences. During these days, he showed inconsistent improvement in using adjectival phrases as compared with his performance during the pre-interventionist interview. He needed three hints in the pre-interventionist interviews, but he fluctuated between one to four hints during the first three days of the DA phase. His conjugation of the past tense also showed improvement when compared to his performance in the pre-interventionist interview by needing only one hint during the third lesson.

After 5 days into the DA phase, he started to produce long utterances with assistance from the teacher-researcher, which is an improvement over his performance during the pre-interventionist interview when he completely avoided producing any long utterances. For example, the teacher-researcher provided the whole class with a buzz lecture (a quick explanation) on how to use two verbs in the same sentence and how to perform better on the ILR scale in general. They learned that the verb following the conjunctives “أن، ل” the first translates to “in order to” and the second renders the

meaning of “to” in English, had to be in the present tense. He reached independent performance of this feature by the end of this phase.

By the end of this phase, he still needed one or two hints in conjugating the past tense, adjectival phrases, and nouns in construct. These features were necessary to control all tenses or to produce long utterances as described for Level 2 in the ILR scale/column 3 (Appendix D). He performed, however, all of these features independently most and not all of the time during the post-interventionist interview as shown in Table 2. This lack of consistency in producing basic grammatical features reflected that his performance met the descriptors of the second column (Appendix D), which reflected that he advanced from meeting the descriptors of column 1 to fulfilling the standards of column 2. In other words, he advanced to a higher performance of 1+ toward Level 2.

Hazem

Table 3

Comparison of the Pre- and Post-Interventionist Hints: Hazem

Language Feature Assisted	Hints		
	Pre-Interventionist	Post-Interventionist	Change
Measure IV	5	0	-0
Verbal Noun	1	0	-1
Active Participle	5	1	-4
Hypothesizing	5	-	0
Measure X	5	4	-1
Measure V	5 (no derivatives)	3 (with derivatives)	-2
Passive Voice	5	0	-5

Note. The dash used in all tables shows that the feature was not used and not avoided. Derivatives in this table mean other forms of the same verbal root.

Hazem's performance in the pre-interventionist interview and the beginning of the DA phase met the standards listed for the second column of Appendix D. Table 3 showed that he needed only one hint to produce a verbal noun and five hints in all the features that were needed for his long utterances to be coherent as required for Level 2, which was reflected in the higher sublevel of column 3 of the Dynamic Assessment Rubrics Form. The other features for which he needed assistance were not frequently used and required the speaker at Level 2 to produce them with minimum cohesiveness. In the first five lessons of the DA phase (10 teaching lessons), Hazem needed only one hint to independently produce some basic grammatical features as described at column 3. These features were starting the sentence with a verb, adjectival phrases, verbal nouns, feminine plural, and negation. His performance of these basic grammatical features while producing long utterances became independent, and he met the descriptors of column 3, Level 2, after 5 days into the DA phase.

In the remaining days of the DA phase, Hazem seldom needed a hint to produce a basic grammatical feature such as conjugating the present tense for which he needed initially one hint only. He started in this remaining period to produce the passive-voice form with the assistance of three hints. By the end of the DA phase, he produced this feature independently, but he needed three hints to produce the passive voice of less frequently-used verbs or measures. He also started to combine features to utter more complex structures such as a noun in construct using a verbal noun needing the assistance of three hints. Example for producing more complex utterance was using “أَنْ” (to) before a present tense and a noun in the singular or plural form after “كُلِّ” (all or each).

Two days after the teacher-researcher gave the buzz lecture (a quick explanation) mentioned above in Basem's section, however, Hazem started to combine two verbs in one sentence using the proper conjunctions “أَنْ ، لَ” and to follow these conjunctions with a present tense. He ventured out and used other less commonly produced verb measures such as the reflexive form of Measure VIII (one of the Arabic 10 verb measures/forms taught) needing three hints only. He started to use the passive voice for different measures using 3 hints from the teacher-researcher. Hazem continued to improve his structural control in Arabic until he performed independently in the post-interventionist interview all the required features for Level 2 and for which he needed assistance during the pre-interventionist interview

Ibrahim

Table 4

Comparison the Pre- and Post-Interventionist Hints: Ibrahim

Language Feature Assisted	Hints		
	Pre-interventionist	Post-interventionist	Change
Noun in Construct	1	0	-1
Relative Clauses	Avoided	0	-5
Adjectival Phrases	1	0	-1
Long Utterances	Avoided	0	-5
Measure III	Avoided	3	-2
Using Present Tense After أَنْ	Avoided	0	-5
Passive Voice	Avoided	0	-5

Ibrahim started the DA phase meeting the descriptors of column 2, because he did not need any assistance with the basic grammatical features as required for Level-2 speakers. He completely avoided producing long utterances, however, and consequently all the features needed for their coherence. The features avoided, as shown in Table 4, were relative clauses, Measure III, using the present tense after “أَنْ” (to) and the passive voice. He started the DA phase needing only one hint for the passive voice that he avoided during the pre-interventionist interview. He needed one hint only to conjugate the present tense, starting a sentence with a verb, adjectival phrases. He started to produce these features independently with no assistance after 4 days in the DA phase. He also independently produced the present tense after “أَنْ” (to) with dropping the plural “نْ” as required after 4 days also into the DA phase.

His independent performance of long utterances moved him to the higher sublevel on Appendix D, column 3/Level 2. From this point until the end of the DA phase, he needed assistance only with less frequently used features. For example, he needed three hints to produce irregular-plural forms or using singular-feminine noun after any number bigger than 10. Not only did he independently produce a present tense in the plural form after “أَنْ” (to), but also he helped Jamal with this feature and consequently Jamal needed only two hints to utter it correctly. He needed one hint also with an irregular use of the gender agreement when he erroneously conjugated a verb in the singular masculine in reference to Egypt. The femininity of countries is very inconsistent in Arabic and difficult for learners to control easily. The features that Ibrahim needed assistance with in the last part of the DA phase were not required for Level 2 and some of them would be controlled

by Level 3 speakers. As reflected in Table 4, Ibrahim independently performed all the required features for Level 2 during the post-interventionist interview.

Jamal

Table 5

Comparison of the Pre- and Post-Interventionist Hints: Jamal

Language Feature Assisted	Hints		
	Pre-interventionist	Post-interventionist	Change
Hollow Verbs	4	-	-0
Conjugating Past Tense	3	0	-3
Gender Agreement	5	0	-5
Long utterances	Avoided	0	-5
Present Tense	4	0	-4
Using Present Tense after أن	5	1	-4
Passive Voice	Avoided	0	-5

Jamal started the DA phase meeting the standards of 1+ as listed in the first column of the Dynamic Assessment Rubrics Form (Appendix D). In the interventionist interview, he needed assistance with basic grammatical features and avoided producing long utterances. Table 5 showed that he needed three hints to conjugate the past tense, five hints to control gender agreement, four hints to conjugate past tense, and five hints to use a present tense after أن (to). He dropped his need for assistance in conjugating the past tense to two hints and to one hint for the present tense. He also needed two hints to produce a verbal noun at the beginning of the DA phase, but 4 days later he needed only one hint. Later he started to use long utterances for which his needs for assistance with

cohesive devices such as “أَنْ” started to surface. He continued the remaining six lessons needing one or two hints still to produce verbal nouns for different measures, conjugating tenses properly. Therefore, Jamal finished the DA phase meeting the descriptors of column 2, which was at a much higher point in the range from Level 1+ to Level 2. The information in Table 5 showed his independent performance during the post-interventionist interview in most of the features for which he needed assistance during the pre-interventionist interview. The other features that he had not developed sufficiently enough for independent performance impeded him from consistently producing long utterances as required for Level 2. Although this inconsistency in producing long utterances prevented him from meeting the standards of Level 2/column 3, his performance still was sufficient to fulfill the criteria of column 2. This advancement to column 2 recorded his advancement from his starting performance of column 1 in the pre-interventionist interview.

Salwa

Table 6

Comparison of the Pre- and Post-Interventionist Hints: Salwa

Language Feature Assisted	Hints		
	Pre-interventionist	Post-interventionist	Change
Verbal Noun	Avoided	4	-1
Conjugating Past Tense	3	0	-2
Adjectival Phrase	4	0	-4
Present Tense	1	0	-1
Using Present Tense after أَنْ	Avoided	0	-5

Salwa started the DA phase at column 1, because she avoided producing long utterances and needed assistance with basic grammatical features as shown in Table 6. She needed four hints on the first day of the DA phase to produce an adjectival phrase. She improved her performance of the adjectival phrase during the first 4 days by needing only one hint. During these first 4 days, she also needed three hints to produce a noun in construct, one hint to conjugate past tense, four hints to appending personal pronouns to a preposition, one hint to refer to a country as a feminine noun, and four hints for irregular plural. During the remaining six lessons, she kept improving until she independently performed these features as shown in Table 6.

Although Salwa started to produce coherent long utterances most of the time on the 6th day of the DA phase, she continued to struggle with several basic grammatical features. For example, she needed two hints to use a possessive pronoun, one hint to produce a verbal noun, and one hint to produce an adjectival phrase. Her longer utterances, however, reflected improvement in using “أن” (to in English) between two verbs. She needed only one hint to produce this feature by the end of the DA phase, although she avoided it completely during the pre-interventionist interview. Her structural control advanced enough to produce a complex form by combining the passive voice with the irregular plural with the assistance of only two hints. She finished the DA phase meeting the standards of column 2 of Appendix D.

*Ramzy***Table 7****Comparison of the Pre- and Post-Interventionist Hints: Ramzy**

Language Feature Assisted	Hints		
	Pre-interventionist	Post-interventionist	Change
Relative Clauses	3	0	-0
Negation in the Past	1	0	-1
Adjectival Phrases	5	1	-1
Verbal Noun	Avoided	2	-3
Conjugating Past Tense	5	1	-4
Using Present Tense after ﺃ	Avoided	3	-2

Ramzy started the DA phase meeting the sublevel of column 1 of Appendix D, because he needed assistance with basic grammatical features during the pre-interventionist interview as shown in Table 7. He needed three hints to negate the past tense, five hints to produce adjectival phrases, and five hints to conjugate the past tense. He also avoided verbal nouns and using the present tense after “أَنْ” (to in English). One advanced feature toward the higher sublevel on Appendix D that he used with 3 hints during the pre-interventionist interview was the relative clause. The use of these forms improved during the first 3 days by using at least one hint less to produce them. He needed only three hints to use the past tense, two hints to convert a noun to an adjective, four hints to produce the adjectival phrase.

He used different forms in the DA phase that did not surface during the pre-interventionist interview. He needed one hint to conjugate verbs of Measure IV, three

hints to construct a passive voice in the future tense, and using “أَنْ” (means that in English) before a nominal sentence. This particular feature is usually used for long utterances. His inconsistent production of long utterances advanced him to the higher sublevel on Appendix D, because he met the descriptors of column 2.

For the last 3 days of the DA phase. He produced the passive voice with only two hints. He used the noun-in-construct with the assistance of two hints; this feature was even avoided during the pre-interventionist interview. He also produced the verbal noun with only one hint, which was a substantial improvement compared to the five hints he needed at the beginning of the DA phase. He used “أَنْ” (to in English) in between two verbs independently without any assistance. Table 7 showed substantial improvement for his performance during the post-interventionist interview as compared to the pre-interventionist interview. This improvement matched his progress during the DA phase.

Ramzy advanced quickly beyond all expectations. He expressed his desire to drop out of this study on the 2nd day of the DA phase feeling that his speaking ability was much weaker than everyone else’s in class. Then, he changed his mind. Ramzy was extremely introverted and quiet and his interactions were not voluntary during the DA phase. Although both the post-interventionist interview and his performance in the last few days of the DA phase showed him advancing to the descriptors of column 2 of the Dynamic Assessment Rubrics Form (Appendix D), Ramzy continued to improve after this study to join Ibrahim and Hazem by achieving Level 2 for the formal OPI.

To investigate further the change that happened from the pre- to the post-interventionist interviews for the whole group, the Arabic features for which the participants needed the teacher-researcher’s assistance had been tallied to calculate the

mean (\bar{x}) for each one of them. Table 8 shows the comparison between the Pre- and post-interventionist means (pre- and post- \bar{x}).

Table 8
Comparison of Language Features Means: Pre- and Post-Interventionist

Language Feature	Recipients	Hints				
		Pre-interventionist	Post-interventionist	Pre \bar{x}	Post \bar{x}	Change
Adjectival Phrase	4	13	1	3.3	.3	-3
Present Tense After أُ or ل	5	25	4	5	.8	-4.2
Conjugating Past Tense	4	14	1	3.5	.3	-3.2
Long Utterance	4	20	0	5	0	-5
Verbal Noun	4	13	6	3.3	1.5	-1.8
Passive Voice	4	20	0	5	0	-5

Note. A negative number in the “Change” column reflected lesser number of hints. Providing lesser number of hints reflected improvement by being closer to independent performance as described for the higher sublevel. Improvements are displayed in boldface.

All the assisted features that were in common for students are shown in Table 8.

The other students who were not factored into Table 8 showed improvement for the same features during the daily interactionist-DA of the DA/TBLI instructions. A comparison between the mean of the number of hints provided to students during the pre-interventionist interviews (pre- \bar{x}) and the mean of the number of hints provided to students during the post-interventionist interviews (post- \bar{x}) is presented in Table 8. The pre- \bar{x} and post- \bar{x} were calculated by averaging the total number of hints provided for a particular feature divided by the number of students who were assisted. The change was stated as a negative number when the post- \bar{x} was smaller than the post- \bar{x} to show that the

number of hints provided was smaller than what it was initially. This negative number reflected improvement in performing the assisted feature and consequently recording an improvement in the students' structural control of Arabic.

All students showed positive change in their performance of all the features shown in Tables 2 to 8, which indicated that all participants improved in the structural control of Arabic. The progress of each student in between the pre- and post-interventionist interviews, however, was examined by reviewing the Dynamic Assessment Rubrics Forms used daily during the DA phase. The teacher-researcher used the designed rubrics (Appendix C or D) in the classroom during the DA phase, and he tracked each student's progress day by day.

Reviewing the pre- and post-static and dynamic interviews and evaluating the progress of the six participants during the DA phase demonstrated a positive change in their structural control of Arabic speaking. All participants improved their structural control of Arabic for many features described in the ILR scale. Comparing the pre- and post-OPIs, the pre- and post-interventionist interviews, and the daily interactionist-DA/TBLI instructions showed an improvement of the structural control of Arabic speaking for all students.

Research Question 2

Research question 2 asked: How do OPI without DA assistance and OPI with DA assistance compare relative to the evaluation of Arabic speaking? Comparing OPI without DA assistance and OPI with DA assistance showed that both of them were capable of evaluating the proficiency level of the examinee, but the OPI with DA assistance was more effective in providing diagnostic feedback on a daily basis or prior to

the DA phase. Both types of assessment were designed to evaluate learners' proficiency levels in general and Arabic in particular in the case of dynamic assessment.

OPI was specifically designed by the U.S. Government to evaluate the speaking proficiency level of a foreign language, that is, OPI was a summative, static, and psychometric test that was designed to rate examinees' proficiency levels of a foreign language. The interventionist DA, however, was designed as a diagnostic and a formative interview. Question 2 of this study prompted the comparison between these two types of assessment. This comparison revealed their strengths and weaknesses in accomplishing each other's function. In other words, it compared the capability of OPI to function as a diagnostic formative instrument and the efficiency of the DA as a summative test. A comparison between the results of both pre-OPIs and pre-interventionist interviews and the post-OPIs and the post interventionist interviews is shown in Tables 9 and 10. This comparison investigated the regular OPI's capability to function as a formative diagnostic test and to explore the possibility of the interventionist interview to function as a summative test.

Table 9**Evaluative Feedback of the Pre-OPIs and Pre-Interventionist Interviews**

Name	Pre-OPI	Pre-interventionist Interviews
Basem	1+	1+
Hazem	High 1+	1+
Ibrahim	Low 1+	1+
Jamal	Low 1+	1+
Salwa	Low 1+	1+
Ramzy	Low 1+	1+

Note. This table shows only the rated proficiency level for each student by the pre-interventionist interview. Refer to table 2 to 7 for the detailed diagnostic information for every student by the same type of assessment.

The results of evaluating the proficiency level of students in the pre-OPIs were 100% the same as the results of the pre-interventionist interviews as shown in Table 9. All students were evaluated to be at proficiency Level 1+ by both types of assessment. There was 100% agreement between these two types of assessment in rating the proficiency level of the six participants. The diagnostic information from Table 4 and the part of Ibrahim's results mentioned above in the section on Question 1 were compared with the information in Table 9. Both types of assessment did not have the same diagnostic results for Ibrahim.

The pre-OPI evaluated his performance as a "low" 1+, whereas the pre-interventionist interview evaluated his performance by the descriptors of the second column of the Dynamic Assessment Rubrics Form, which was the closest detectable performance to Level 2 (Appendix D). Considering that testers usually used the jargon "low" to mean the lowest performance of a given proficiency level, both types differed

on Ibrahim's strength of performance in the range of abilities between 1+ and 2. The OPI testers described it by the unidentified "low," and the interventionist-DA interview specifically described his performance by the descriptors of column 2 and the specific number of hints he needed for every undeveloped feature.

Even with this discrepancy between the two types of assessment on diagnosing this student, the percentage rate of agreement between them could be described as 83%, that is, considering that the testers unidentified "low" is equivalent to column 1 and "high" as congruent to column 2 of the Dynamic Assessment Rubrics Form (Appendix D). There was a parallel reliability coefficient of .80 between these two types of diagnostic information. The data in Tables 2 to 7, however, show that the pre-interventionist interviews provided much more diagnostic details for each student. The detailed diagnostic information was more accurate and measurable than what the regular OPIs provided as a feedback. The OPI testers described the examinee's performance with the undefined descriptor of "low" or "high." The pre-interventionist interviews, however, provided the language features for which each student needed assistance and how far each was from independent performance as described for the higher sublevel on the Dynamic Assessment Rubrics Form.

Table 10**Evaluative Feedback between Post-OPIs and Post-Interventionist Interviews**

Name	OPI 2	Post-Interventionist
Ramzy	1+	1+
Salwa	1+	1+
Ibrahim	2	2
Jamal	High 1+	1+
Basem	1+	1+
Hazem	2	2

Note. This table shows only the rated proficiency level for each student by the pre-interventionist interview. Refer to Tables 2 to 7 for the detailed diagnostic information for every student.

Both types of assessment agreed 100% on rating the proficiency level of all students as it is shown in Table 10, which meant that there was a parallel coefficient of 1.0 between these two types of assessment. As far as comparing the diagnostic feedback, only one student's performance, Jamal, was described as being "high" 1+ by the post-OPI. The post-interventionist interview agreed with this assessment as it is shown in Table 5 and in the part on Jamal's performance during the DA phase in the section of Question 1 above. This agreement on diagnosing Jamal's performance was based on the assumption that the testers' unidentified "high" is equivalent to the descriptors of column 2 of Appendix D. The other testers did not volunteer to give any further description for the other students' performances, which is understood usually to mean the average performance of 1+. Tables 2 to 7, however, showed that the post-interventionist interviews provided more detailed diagnostic data for each student than any informal OPI. The post-interventionist interviews provided feedback data that were measurable,

detectable, and based on the ILR scale. Furthermore, the feedback back data showed the potential learning for each student by showing features that still needed assistance for the learner's performance to meet independently the higher sublevel as described on the Dynamic Assessment Rubrics Form.

Students expressed their opinions about the diagnostic ability of the DA interviews or the DA/TBLI instruction in two data sources. One source was the second item of the ten 5-point scales. This item elicited the students' responses on the following statement: DA/TBLI instruction was capable of diagnosing each student's language needs on a daily basis. Three students responded by marking "agree," and three checked "strongly agree" on this statement, that is, all students agreed that DA/TBLI instruction was capable of diagnosing each student's needs on a daily basis.

The other data source was during the interviews conducted by the teacher-researcher after the post-interventionist interview for each student. Their responses indicated that the daily DA/TBLI process was capable of diagnosing students' needs accurately. The students' responses during the interview, the ten 5-point scales, and the information reviewed from Tables 2 to 10 showed that both OPI without DA assistance and OPI with DA assistance were capable of evaluating students' proficiency levels by the ILR scale. The OPI with DA assistance, however, was the only one capable of providing detailed diagnostic information that was accurately detectable and calibrated by the ILR scale.

Research Question 3

The third question of this study asked: How do the experiences and perceptions of DA/TBLI instruction compare between teacher-researcher and OPI testers? Both observers and the teacher-researcher agreed on the following themes as determined by analyzing data from interviewing the observers and from the teacher-researcher journal. The first question for interviewing the observers was: what is your perception about the diagnostic abilities of the DA/TBLI instruction? Nine out of 10 observers were interviewed immediately after the lesson they observed. All of them agreed that the daily DA/TBLI process was capable of diagnosing students' needs accurately. One observer declared, "it is very helpful in diagnosing accurately students' needs." Another observer stated "Focusing on the structural control made the process very effective, and especially that dynamic assessment and task-based language instruction went well together in the classroom."

Observers expressed also the following thoughts in their responses to the first question of their interviews. The language features were diagnosed in a real-life context, and, therefore, the teacher was capable of diagnosing their form, meaning, and use as described on the ILR scale. One observer expressed that the fact that the feedback for the identified deficiency was instantaneous elevated the students' focus on the assisted feature during the hinting process. Two observers, however, expressed concern about raising the students' affective filter (Krashen, 1987) that might harm the students' fluency for the sake of accuracy. Fluency was the landmark for the ILR's proficiency Level 2 in general and the accuracy factor of delivery in particular.

The second question in the interview was: How practical is using the DA rubrics in class while teaching? Their responses indicated unanimously that using the rubrics of the daily DA/TBLI instruction was practical. The percentage rate of agreement on the practicality of the DA/TBLI instruction was 100%. All observers, however, agreed that the practicality could increase even further by simplifying the Dynamic Assessment Rubrics Form. One observer said, “This method will be more practical, if the form became simpler.” They all agreed that observing the teacher-researcher filling out the form while the students were busy in their small groups did not take away from the teaching requirements of the class. One of the observers told the teacher-researcher, “I noticed you entered your observation quickly in the form while the students were busy in their small groups.” The teacher-researcher noticed that carrying the form with him on a clipboard facilitated the process very much, because it gave him the opportunity to enter seamlessly the number of hints provided once an opportune moment became available. One observer also suggested video or audio recording the lesson to double check or to supplement later the entries made in the classroom. This observer shared, “I think this method would be more practical, if you recorded this lesson to further your entries in the form later in your office.”

Additionally, and based on the interviews done with two observers on the first 2 days of the DA phase, the teacher-researcher took notes of their suggestions to simplify the Dynamic Assessment Rubrics Form. He actually continued this process until the last week of instruction. In summary, all observers agreed that using the DA/TBLI rubrics were practical to use while teaching during their observed lessons.

The fourth question in the interview was: Do you think teachers need training on using the Dynamic Assessment Rubrics Form before using it in classrooms? They all agreed that teachers should be trained on the DA/TBLI instruction and its rubrics before implementing it in classrooms. One observer mentioned “They need training and observation of others using it, and to be normed on interpreting the Dynamic Assessment Rubrics Form.” Norming was an expression used in DLIFLC to indicate having common understanding of terms and interpretation of the ILR standards. Two observers suggested hands-on training on designing task-based lessons that would be suitable for the students’ diagnosed needs. Observers expressed that teachers would need to develop the skill of selecting suitable material for the students’ proficiency level and needs to plan a time-efficient lesson. One observer announced, “Of course, training teachers will be needed to create a lesson plan that is time efficient. The teacher needs training on using it and on using the hinting process.” All observers expressed also that instructors would need experiential training that would include peer-observation, because they noticed that DA/TBLI made a difference. DA/TBLI instruction making a difference was the following theme that was found in both transcriptions of the interviews and the teacher journal of the teacher-researcher.

This theme emerged out of the interview question about if the observers thought that DA/TBLI made a difference. Their responses indicated that DA/TBLI instruction made a difference by giving students a chance to reflect on their performance to realize the missing parts in their knowledgebase and skills. One hundred percent of the participating observers and the teacher-researcher agreed that this approach made a difference, because it enhanced the students’ involvement with each other, with the

teacher-researcher, and reflecting on their own performance. Their responses indicated that students' involvement with each other and the teacher-researcher enhanced their learning through their exchange of knowledge and ideas. The same involvement prompted their reflection on their own performance to realize what they missed through the hinting process.

Four observers expressed the students' enthusiasm in different ways. The first observer reported that he noticed they were more enthusiastic to learn because DA/TBLI instruction made them more focused and engaged. The second observer mentioned that students did not mind the nonintrusive way of the teacher-researcher to the extent that students welcomed the hinting process. The third observer mentioned that students were happy to realize that they knew a part of the needed knowledge or the skill to perform a language feature independently. The fourth observer noticed the students' enthusiasm also when he observed them on their way out after the lesson comparing the number of hints needed for the different language features assisted. This point led to the following theme about the observers' and the teach-researchers' level of enthusiasm.

Observers were enthusiastic about the future implementation of the DA/TBLI instruction. One hundred percent of all interviewed observers expressed their admiration of the DA/TBLI instruction and contributed several suggestions for its improvement or for its future implementation in DLIFLC. Their suggestions are discussed in the next chapter; however, they are only listed in the remaining part of this paragraph. Three of the observers suggested ideas for enhancing the practicality of Dynamic Assessment Rubrics Form. Three observers suggested ideas for increasing the time available for student-student and student-teacher interaction. One observer suggested that the

homework for the previous day could be designed to prepare students for the lesson so that the teacher could save the time of the pretask activities. Another suggestion was allocating 2-hour blocks of instruction for the DA/TBLI lessons. The observers' enthusiasm was shared by the students. The next section presents the findings synthesized from interviewing the students and from their responses to the administered survey.

Research Question 4

The fourth question of this study was: What are student perceptions of the DA process? Student perceptions were very positive about DA/TBLI instruction's capability for diagnosing their language needs daily and for promoting learning. General themes were determined from two data sources. These data sources were from interviewing each student after the post-interventionist interview and from the student survey of the ten 5-point scales. The response to this survey was anonymous online and per hard copy.

The following are the results of the ten 5-point scales. This survey included nine statements, and students were asked to respond to each one by selecting one of five options: *strongly disagree*, *disagree*, *I do not mind it/similar to regular instruction*, *agree*, or *strongly disagree*. These five options corresponded to values graduated from one to five points in the same order. The last item was an open-ended item for students to enter their additional comments or remarks. Only one student responded to this last item. Table 11 shows the responses to nine of the ten 5-point scales.

Table 11
Frequency of Responses to the Ten 5-Point Scales

Scales	Strongly Disagree	Disagree	I do not mind it	Agree	Strongly Agree
1. The DA/TBLI instruction method is an effective classroom approach for language learning	0	0	0	3	3
2. DA/TBLI instruction is capable of diagnosing each student's language needs on a daily basis.	0	0	0	3	3
3. The hinting process helped me overcome my personal language difficulties.	0	0	0	4	2
4. The hinting process that I experienced improved my speaking ability in Arabic quickly.	0	0	0	3	3
5. I would recommend DA/TBLI instruction for other language students.	0	0	0	2	4
6. Knowing the ILR standards helped me understand what I need to do to improve my speaking abilities.	0	0	0	4	2
7. Collaborating with other students to deliver a measurable product provided me with a great learning environment.	0	0	1	1	4
8. Following other students going through the hinting process helped me learning and or overcoming my own personal difficulties.	0	0	0	5	1
9. Using DA/TBLI instruction in the classroom was practical and enjoyable.	0	0	0	4	2

The frequencies in Table 11 show that the students' responses to the first nine scales were either *agree* or *strongly agree*. One exception to this statement is found in their responses to scale 7 for which only one student selected *I don't mind it*. All the statements used in scales 1 to 9 are positive statements about the DA/TBLI instruction and all its relevant topics and activities. The results for the 5-point scales as shown in Table 11 supported the themes found by interviewing students individually.

Only one student responded to the open-ended item 10. This student wrote, "Strongly recommend this program be enacted in DLI at a minimum during the speaking hour. The only difficulty in obtaining it would be due to a lack of teachers knowing the subject or if the teacher was lacking in language skills him/herself. Also, would recommend that the teachers NOT pervert this into simply a reiteration of OPI topics over and over and over again as that creates boredom, extreme boredom and will cause the students to put forth very little effort. Topics do not always have to be Middle East focused, just keep the students interested and the conversation will flow, allowing the teacher to do nothing but pay attention to mistakes and hint when necessary. Knowing the topic well would be great to have lively discussions which would greatly increase student participation and learning in the topic and more importantly in the language."

Four themes for the students' perception about the DA/TBLI approach were found by interviewing students and from conducting the ten 5-point scales as shown above. One, the daily DA/TBLI instruction was capable of diagnosing students' needs accurately. The five students interviewed thought that the DA/TBLI instruction was an accurate tool of diagnosing students' needs daily. One student said: "It helped to know

what I didn't know" Moreover, students supported this theme further in their responses to the second item of the ten 5-point scales.

Two, DA/TBLI made a difference in the students' learning/performance of Arabic speaking. The five students available for the interviews agreed that the DA/TBLI instruction enhanced their learning process. Their responses included that this approach increased their wealth of vocabulary and their retention of the new vocabulary items and newly developed language structures. For example, one student mentioned, "I felt this approach improved my listening ability." Another student stated, "It increased my engagement during the lesson." One other response indicated that the student was able to overcome his language difficulties quickly. Moreover, 100% responded positively to the statements in the scales except scale 7. Only one response was neutral for the statement in scale 7 on the collaboration with other students.

Three, concurrent and cumulative techniques of DA promoted learning effectively. All students interviewed expressed their agreement on benefitting from the hinting process with other students in the class. One response elaborated on this issue by declaring, "Half the errors produced were shared by all students, and we would learn the correct utterance when one of us went through it with you." The students' responses to scale 8 that declared a positive statement following others going through the hinting process in the classroom supported this theme. Five students agreed and one student strongly agreed, which meant that 100% of the participants agree that the concurrent and the cumulative techniques of DA promoted learning in the classroom. Responses to the statement of scale 7 reflected four students strongly agreed, one agreed, and only one did not mind. That is, 83% of the participants agreed that the necessary venue for the

concurrent and the cumulative techniques of DA provided a great learning environment, whereas only one student out of six participants chose “I don’t mind it” about collaborating with others.

Four, selecting the material was crucial for the DA/TBLI approach. Students expressed that having interesting material and activities were pivotal to the effectiveness of the hinting process. Students expressed that enjoying the topic and the material was necessary for their collaboration and engagement with the material and consequently the hinting process. The one response to item 10 cited above supported this theme. Students mentioned in their interviews also that when the reading or listening material was too challenging, they lost some of the time available for speaking and collaborating. One student said: “I spent too much time trying to process the material when it was difficult” Consequently, it diminished the time allocated for the hinting process.

Summary

This chapter presented the findings for the study’s four questions. The findings for the study’s first question reflected a positive change in the structural control of Arabic speaking based on the DA/TBLI instruction. Five of the six participants improved their structural control of several Arabic features in the post-OPIs and 100% showed improvement in the post-interventionist interviews. The findings of the second question indicated that both OPI without DA assistance and OPI with the DA assistance were capable of evaluating a student’s proficiency level, but only the OPI with DA assistance and the interactionist-DA were capable of diagnosing accurately and measurably students’ needs to advance on the Dynamic Assessment Rubrics Form and the ILR scale.

There was a 100% agreement between both the pre- and post-OPIs and interventionist interviews in evaluating the students' proficiency. The results for both types of assessment awarded the same proficiency level for all students in both iterations. The OPI with DA assistance (interventionist DA) and the DA/TBLI instruction (interactionist DA) demonstrated the capability of providing detailed and accurate diagnostic feedback based on the ILR scale. One hundred percent of all students and the observers agreed that dynamic assessment was capable of providing accurate diagnostic feedback daily.

The synthesis of questions 3 and four is presented in this next part of the summary. One hundred percent of all observers and students agreed that DA/TBLI instruction was practical on daily basis, and it increased students' involvement and enthusiasm during the lesson. All students and the observers agreed that the DA/TBLI instruction was an effective approach of teaching Arabic speaking in classrooms. On the one hand, all observers, however, agreed that teachers would need training on designing DA/TBLI instruction, the Dynamic Assessment Rubrics Form, and experientially filling out the Dynamic Assessment Rubrics Form in class. On the other hand, students supported this theme further by their responses to scale 1 on the ten 5-point scales. All students agreed in their responses to the scale stating that the DA/TBLI instruction method was an effective classroom approach for language learning.

The following summarizes the results of the ten 5-point scales. One hundred percent of the students thought that the hinting process helped them overcome their personal language difficulties and improved their speaking ability in Arabic. All students would recommend DA/TBLI instruction for other language students. One hundred

percent of the students agreed that knowing the ILR standards helped them understand what they needed to do to improve their speaking ability. Five out of six students agreed that collaborating with other students to deliver a measurable product provided them with a great learning environment. Only one student expressed a neutral attitude about collaboration with others in class. All students agreed that following other students going through the hinting process helped them learn and or overcome their personal language learning difficulties. Last, all students agreed that DA/TBLI instruction in the classroom was practical and enjoyable.

CHAPTER V

SUMMARY, DISCUSSION, AND CONCLUSION IMPLICATIONS

This chapter is divided into four parts. The first part includes a summary of this study. This summary includes the background, problem, purpose, and the questions of this study. The second part contains a discussion of the findings of this study's questions. The discussion of each question is presented in a subsection in which an interpretation of the findings is given. The third part offers a synthesis that provides a broader understanding of the DA/TBLI approach in teaching Arabic. The fourth part presents the implications of this study. These implications are divided into recommendations for practice and for future research.

Summary of Study

Previous studies (Hill & Sabet, 2009; Lantolf & Poehner, 2011; Poehner, 2005; Poehner & Lantolf, 2005) on dynamic assessment failed to address one or more of the following elements: (a) the use of dynamic assessment in language classrooms, (b) the instructional activities used with adult learners, (c) the input materials used with students, and (d) the scale on which the dynamic-assessment process was calibrated. Dynamic assessment was used mainly in a tutoring format, and the studies using it in a classroom setting did not mention the method of teaching used with adult learners. Other studies in a classroom setting were conducted with children (Lantolf & Poehner, 2011). No studies elaborated on the rubrics used to evaluate the daily progress of students. They showed students' progress by comparing certain language features before and after an enrichment program. These features were grammatical structures that were not associated with any standardized language scale.

This present study combined dynamic assessment and task-based language instruction in the planning and implementation of its daily lessons. Task-based language instruction (TBLI) met all the principles of adult learning and it prompted students to use their target language meaningfully. Meaningfulness meant that students used the language in a real-life scenario using authentic material. The authentic materials were selected for this study by using the principles of text typology and the Interagency Language Roundtable (ILR) scale. They were selected also according to the students' personal interests and current abilities of their reading and listening skills in Arabic. The authentic material used during the DA phase of this study prompted students with the context in which they needed to solve a real-life situation. Their solution had to be in the form of a measurable language outcome that could be referred to in this study also as a measurable language product.

During their involvement in the assigned task in their small groups of two or three students to generate this product, the teacher-researcher found many opportunities to use dynamic assessment concurrently or cumulatively (Hill and Sabet, 2009). The teacher-researcher provided the gradual hints of dynamic assessment selectively to help students improve the control of the structural features of Arabic that had not developed completely. These features were referred to as immature in the context of DA in general and in this chapter in particular. He recorded the number of hints provided on the Dynamic Assessment Rubrics Form (DARF), which showed their current proficiency level in speaking Arabic and the number of hints provided. The teacher-researcher devised and developed this form by deconstructing the proficiency levels of the ILR scale into detectable and noticeable sublevels.

The purpose of this study was to investigate the effectiveness of combining dynamic assessment with task-based activities that would target the speaking skill of Arabic. It investigated the practicality of continually assessing students' weaknesses and strengths during their course of instruction and particularly as a group (Brown, 2009; Ellis, 2009a). This research was designed to use the Interagency Language Roundtable (ILR) scale, which was the proficiency scale used in the U.S. Government with students attending the Defense Language Institute Foreign Language Center. Studying the effect of DA/TBLI instruction was measured by using the Interagency-Language-Roundtable rubrics to see the change in students' performance at the end of this study. To make this measuring more practical for the purpose of this study, the focus was only on one accuracy factor of the proficiency levels of the Interagency Language Roundtable scale. The accuracy factor measured in this study was Arabic "structural control." To measure the effectiveness of the DA/TBLI approach on adult learners of Arabic, this study addressed the following research questions:

1. What is the change in the structural control of Arabic speaking based on DA/TBLI instruction?
2. How do OPI without DA assistance and OPI with DA assistance compare relative to the evaluation of Arabic speaking?
3. How do the experiences and perceptions of DA/TBLI instruction compare between teacher-researcher and OPI testers?
4. What are the student perceptions of the DA process?

To answer these questions, the study was designed in three stages: the pre-DA phase, the DA phase, and the post-DA phase. Six students were selected during the pre-

DA phase to continue the DA and the post-DA phases. Students went through the pre-OPI and the pre-interventionist interview during the pre-DA phase. The teacher-researcher used the DA/TBLI approach during the DA phase to teach the participants for one hour daily, and both he and the observers, who were certified OPI testers, used the Dynamic Assessment Rubrics Form. The teacher-researcher interviewed each observer immediately after the daily lessons of the DA phase. Then at the end and during the post-DA phase, students went through the post-OPI and the post-interventionist interview. In addition to interviewing each student after the post-interventionist interview, students responded to a survey of ten 5-point scales. The following part of this chapter is a discussion of the findings to the study's questions.

Discussion

The discussion part of this chapter is divided into four subsections. Each subsection is designated for one of the studies four questions to present a summary of its findings and the interpretations of these findings. These interpretations would be discussed through the lens of the theoretical framework mentioned earlier in the first chapter of this study. The theoretical models of this study were the sociocultural theory (Vygotskey, 1078) and task-based language instruction as a suitable approach for adult learners (Ellis, 2009a, 2009b; Foster & Skehan, 1999; M. H. Long, 2000; Skehan, 1998; Skehan & Foster, 1999).

Question 1

The first question of this study was: What is the change in the structural control of Arabic speaking based on DA/TBLI instruction? Comparing the pre- and post-OPIs, the pre- and post-interventionist interviews, and the daily DA/TBLI instruction showed an

improvement of the structural control of Arabic speaking for all students. Five out of six students, 83% of participants, showed improvement in their proficiency level by comparing the pre-OPIs with the post-OPIs. All students showed improvement in their structural control of Arabic by comparing the results of the pre- and post-interventionist interviews. This success of combining dynamic assessment with task-based language instruction suggests that both approaches would complement each other to maximize second language acquisition for adult learners. The first was a constructivist approach that would empower learners by engaging a stronger peer or a teacher to obtain the information missing in their knowledgebase (Brown, 2009; Doolittle, 1997; Hill & Sabet, 2009; Lantolf & Poehner, 2011; Poehner, 2005; Poehner & Lantolf, 2005; Vygotsky, 1978).

The engagement with a stronger peer or the teacher-researcher created Vygotsky's (1978) zone of proximal development (ZPD). The gradual hints provided enabled the teacher-researcher to identify the borders of the ZPD for every learner or group of learners in the classroom. This area was the range bounded between the student's assisted and independent performances (Brown, 2009; Doolittle, 1997; Hill & Sabet, 2009; Lantolf & Poehner, 2011; Poehner & Lantolf, 2005). The gradual explicitness of the hints provided presented calibrated help for the learner's progress toward his or her independent performance of a certain language feature. Considering that the graduation of hinting was standardized, the teacher-researcher knew how far the student was from performing desirably at the next sublevel of Dynamic Assessment Rubrics Form (Appendix D). This meant that the independent performance of a language feature was

defined by the description for the targeted sublevel on the Dynamic Assessment Rubrics Form.

The hints promoted the quick learning of a language feature, because the teacher or the stronger peer provided the needed incremental knowledge when the student needed it the most (Poehner, 2005). These heightened occurrences of need evolved naturally not only to raise the student's awareness to what was missing in their knowledgebase but also to sharpen their focus while honing in on the teacher-researcher's utterance for its proper performance. This heightened focus while being subjected to the acceptable performance of the language feature might be the reason that led to the student's autonomy soon after (van Lier, 1996) the DA phase or later in the interventionist-DA interviews. The autonomy was defined in this context by the student's independent performance of a certain language feature as described for the higher sublevel on the Dynamic Assessment Rubrics Form (Appendix D). This autonomy was achieved due to the deeper internalization of the incremental knowledge that was added to the student's knowledgebase effectively (Poehner, 2005; van Lier, 1996). This autonomous performance of features needed on the ILR elevated the students' level of motivation knowing that they were improving on the scale by which their language proficiency would be evaluated during their career in the military. For example, students noticed their own improvement toward their autonomous performance by realizing their needs for lesser number of hints to perform the same feature as the DA phase progressed. This realization of the practical work they did in class raised their intrinsic motivation as adult learners (M. H. Long, 2000; van Lier, 1996) who wanted to succeed in their career.

Adult learners would be practical and they would be more motivated when they knew that what they would do in class would help them in their lives (M. H. Long, 2000). Participants knew that the Dynamic Assessment Rubrics Form used was based on the ILR scale, which was the instrument of evaluating their performance for the formal OPI to exit the Arabic Basic Course. Students were motivated in class not only because the hints were provided when their awareness of their importance was the highest, but also because they realized that the process would help them directly perform better in their exit OPI. Based on the students' responses to the survey and the interviews, the social setting provided by the task-based lessons could have made the hinting process successful. In other words, DA/TBLI instruction could be the most suitable mathemagenic venue for using the DA hinting process.

The DA process needed a social setting because it was based on Vygotsky's (1978) sociocultural theory. Task-based language instruction provided the social setting where the need for providing the gradual hints would emerge naturally and meaningfully. The adult participants' realization that they needed assistance to perform in a real-life situation raised their interest in reaching autonomy for the assisted features. This real-life scenario was made possible by following the principles of TBLI (Ellis, 2009b; Nunan, 2004). The relevancy of the introduced material to reality was not the only factor that made TBLI suitable for the adult participants. Their collaboration with each other allowed them to use their previous knowledge to generate the assigned real-life language output (Dean, 2004; Dewey, 1963). This collaboration as adults could be the reason for raising their enthusiasm and preparing the groundwork for creating several ZPDs in the classroom. Group-ZPDs (Hill and Sabet, 2009) were created in each small group, a

teacher-student ZPD, a teacher-group ZPD, and a ZPD between the teacher and the whole class sometimes.

These ZPDs provided a natural environment for the hinting process in cumulative and concurrent settings as it was reviewed and defined from Hill and Sabet's (2009) in the second chapter of this study. The different ZPDs created and these two ways of providing the DA process addressed the diversity of students' proficiency levels and needs. Although all students started this study at proficiency level "1+," the pre-interventionist interview showed that they started at different sublevels of the "1+" range on the Dynamic Assessment Rubrics Form. This diversity could have been the factor that helped the participants during their collaborations in the small groups, because some students were stronger than others in performing certain language features. Consequently the differences in their abilities may have created the desirable ZPDs. Students volunteered to provide each other with hints during their group work and discussing their progress by comparing the number of hints needed after each lesson.

Task-based language instruction may have helped students who came to class with different intellectual styles. Working in all the different modes of Arabic and having to generate authentic outcome yet having to hone in on the teacher-researcher while providing assistance could have addressed the students' preferences of both Type I and Type II intellectual styles (Zhang & Sternberg, 2005). Students' personal background, interests, and abilities in both listening and reading skills were considered in selecting the input material and in distributing them to their small groups in every lesson plan. Samples of these lesson plans are included in the Appendices of this study (Appendix E).

In summary, DA/TBLI instruction was practical, successful, and most likely the main factor that caused improvement in the students' structural control of Arabic. This improvement could be due to the fact that both approaches complemented one another. Both were suitable for the principles of adult learners and students saw progress daily.

Question 2

Question 2 of this study was: How do OPI without DA assistance and OPI with DA assistance compare relative to the evaluation of Arabic speaking? Comparing the pre- and post-OPIs (OPI without DA assistance) with the pre- and post-interventionist interviews (OPI with DA assistance) showed that they had agreement of 1.0 in evaluating students' proficiency levels and .80 agreement in diagnosing students. The same comparison showed clearly that the OPI with DA assistance was far more capable of accurately diagnosing students' needs to advance to the targeted proficiency level on the Dynamic Assessment Rubrics Form, that is, the OPI with DA assistance was the only one of these two types of evaluation capable of identifying accurately, reliably, and with high-validity students' needs.

The high parallel reliability coefficient between both types of OPI in evaluating students' proficiency levels was not a surprise because both OPI and the Dynamic Assessment Rubrics Form were based on the ILR scale. The high parallel reliability coefficient of 1.0 could reflect a high degree of agreement between the descriptors of the sublevels created on the Dynamic Assessment Rubrics Form and the ILR's proficiency levels. During the DA-interventionist interviews (OPI with DA assistance), the teacher-researcher used the same tasks of the OPI, but he also provided the gradual hints of dynamic assessment. Unlike the OPI structure, the teacher-researcher did not follow the

OPI's probing technique to establish the ceiling for the examinee. As explained in chapter III, the ceiling was the term used to indicate the examinee's inabilities in speaking the target language. The teacher-researcher considered the assistance provided to students the ceiling for their speaking abilities.

The high parallel reliability between these two types in evaluating the learner's proficiency levels did not mean that the interventionist interview was a good replacement for the OPI process as a summative test. It might be possible, however, to fine-tune the interventionist DA interviews into a summative test. A confirmation for accepting the provided assistance as the ceiling of the examinee's abilities should be reached statistically first. This statistical research would be imperative before using the DA as a summative test and to make the findings of this study generalizable. Currently and without the statistical due process for validating DA as a summative test, the process of the OPI with DA assistance could not replace the OPI as an instrument for summative evaluation. A different approach might be using tasks prescribed for the higher proficiency level as a probe as was the case in OPI without DA assistance. The interview would need longer time than what a regular OPI would need. It would not be too long, however, and students' fatigue might be mitigated by the friendly assistance provided.

The high coefficient of 1.0 between OPIs with and without DA assistance in evaluating the students' proficiency levels was due to the teacher-researcher's extensive experience in both types. This experience meant that teachers would need to be well-trained as OPI testers and on conducting OPIs with DA assistance using the Dynamic Assessment Rubrics Form. If teachers were trained well on using this approach, they would be able to diagnose students daily in classrooms as well. Teaching diagnostically

by combining dynamic assessment with task-based language instruction, therefore, could be cost effective from a language-program-management point of view. The fact that it would concentrate the efforts of both teachers and students to advance on the ILR scale might be the path to reduce the attrition rate while increasing the number of students accomplishing the objectives of the different courses at the Defense Language Institute Foreign Language Center (DLIFLC).

In this study, DA/TBLI instruction promoted learning while being capable of diagnosing students' needs daily as supported by the findings of question 1 above. At the same time, this cost effectiveness could be furthered if the interventionist DA would prove statistically to be valid and reliable as a summative test in the future. Until that would be accomplished statistically, the OPI with DA assistance (the interventionist interview) would be safer to use as a formative evaluation for diagnosing students' needs and the interactionist DA combined with TBLI would be effective also as a classroom approach for second language acquisition.

The parallel reliability coefficient between these two types of OPI as a diagnostic tool (.8) was not as high as it is for evaluating the proficiency level (1.0). The reason for the drop in the parallel reliability in this case was due to the regular OPI's inability of diagnosing accurately students' needs. The undefined terms of "low" and "high" were meaningless and not based on any specific criteria, which raised serious questions about their interrater reliability. The interventionist-DA interviews, however, were based on validated rubrics that could easily identify the immature language features for the learner. Not only was the interactionist DA able to identify deficient language features, but it also

was able to measure their distances from the independent performance as described for the targeted standard in the Dynamic Assessment Rubrics Form.

Ibrahim was the only student who was diagnosed incongruently between the two types of the OPI as shown by comparing Tables 4 and 9. The pre-OPI diagnosed him as “low” while the pre-interventionist interview diagnosed him as close to Level 2 as shown on Table 4. The daily DA/TBLI instruction diagnosed him as close to Level 2 on the Dynamic Assessment Rubrics Form by evaluating him at the beginning of the DA phase as meeting the descriptors of the second column. Column 2 in the Dynamic Assessment Rubrics Form describes specifically the highest detectable proficiency in the range between Level 1+ to Level 2. The interventionist interview was used to diagnose the learners’ needs at the beginning and at the end of the DA phase as shown in Tables from 2 to 7, whereas the interactionist DA was combined with TBLI not only to diagnose daily the students’ needs but also to promote learning.

The Dynamic Assessment Rubrics Form was used daily to track the hints provided during the DA phase and under the proper column for students’ proficiency sublevels. Ibrahim’s performance met the descriptors of column 2 at the beginning of the DA phase, but he finished the DA phase fulfilling the standards of Column 3. For example, he avoided to produce long utterances, relative clauses, using “أَنْ” (to) in between two verbs, and passive voices in the pre-interventionist interview. These features were necessary as described in column 3 (Appendix D) to meet the standards of Level 2. Column 3 in this form was lifted faithfully from the OPI rating form used in DLIFLC, and the content of column 3 showed the standards of proficiency Level 2 of the ILR

scale. As discussed in the previous chapter, Ibrahim started to use these features independently 5 days into the DA phase.

Accordingly, the two types of DA diagnosed Ibrahim more accurately in the pre-interventionist interview and during the DA/TBLI instruction of the DA phase. To clarify, Ibrahim was one of two students evaluated at Level 2 by both the post-OPI and post-interventionist interview. Considering that the descriptors of column 2 were closer to those of column 3 (Level 2) than the pre-OPI description of “low” (assumingly column 1), then the pre-interventionist interview diagnosed Ibrahim more accurately than the pre-OPI’s description.

Both types of DA evaluated Ibrahim’s diagnostic information more accurately and in more detail than what the regular OPI was able to accomplish. The decrease of the coefficient from 1.0 to .8 was due the precision of the DA process in diagnosing Ibrahim’s needs. To conclude, both types of OPI could evaluate the learners’ proficiency level, but only the OPI with DA assistance was capable of diagnosing accurately and in detail the students’ needs at the beginning and at the end. The interactionist DA was capable of diagnosing the students’ needs daily and most likely promoted the improvement of the students’ structural control.

Question 3

The third question of this study was: How do the experiences and perceptions of DA/TBLI instruction compare between teacher-researcher and OPI testers? Interviews with the observers (10 certified OPI testers) revealed their agreement with the teacher-researcher on the following themes about DA/TBLI instruction: it could diagnose students’ needs accurately, it could be used practically, it could have made a difference in

the students' learning, teachers would need to be trained on it before using, and it should be done for more lessons every day. These themes confirmed and strongly supported the findings discussed above for the study's first two questions. They all agreed that the DA/TBLI instruction was capable of diagnosing accurately students' incomplete (undeveloped) abilities while promoting learning. All observers were experienced OPI testers who had intimate understanding of the OPI proficiency level. Their view that DA/TBLI instruction was a sound approach to diagnosing students' strengths and weaknesses reflected the accuracy and the practicality of the sublevels created for the Dynamic Assessment Rubrics Form. Combining the hinting process with these detectable sublevels fine-tuned the precision of diagnosing students' needs and consequently their potential learning for the subsequent lesson planning as in the results reported from interviewing students, the observers, and the students' survey.

The daily process of diagnosing while teaching a lesson of TBLI allowed the measuring of progress accurately, because the daily forms enabled the teacher-researcher to track the decrease of the number of hints needed for the immature Arabic structures. A trained assessor could realize the advancement of any learner's performance from one sublevel to the higher. Consequently, being trained assessors would prevent teachers from overcorrecting students purposelessly. The practice of overcorrecting might harm the students' fluency and raises their affective filter (Krashen, 1981). Raising the affective filter was a point raised by two of the observers. The teacher-researcher agreed with this point completely and believed that the hinting process should be done selectively for this reason. Selectively means in this context that teachers should only engage students with the hinting process for features needed for their advancement to the

higher sublevel. Additionally, the teacher-researcher selected a language feature that was in common for the whole group or the whole class at the time and he did not deploy the hinting process for too many features simultaneously.

Being selective in deploying the hinting process increased the practicality of the DA/TBLI approach. All observers agreed with the teacher-researcher that DA/TBLI instruction was practical to use in the classroom during the lessons of the DA phase. The reason was, as noticed by nine observers, that the teacher-researcher entered the number of hints provided to students on the form when they were preoccupied in their work groups. He carried the form on a clipboard around the class. The teacher-researcher started to simplify the Dynamic Assessment Rubrics Form by incorporating suggestions from observers starting at the very beginning of the DA phase. The simplest and most practical version of it is shown in Appendix F. By the end of this study, the teacher-researcher had improved his shorthand writing on the Dynamic Assessment Rubrics Form to a much more efficient and practical level. Developing this new skill helped him enter information in the form during the lesson seamlessly and effortlessly.

If the teacher-researcher improved by practicing daily in class, then teachers should be able to be trained efficiently on DA/TBLI instruction. Teacher training was another main theme that was agreed on by all observers and the teacher-researcher. This training should be experiential through classroom teaching, peer observation, and filling out the Dynamic Assessment Rubrics Form while watching video clips of a model lessons. Teachers might come out of this training preferring the technique of filling out the form in the office after teaching a lesson. Teachers could listen also to a recording of the lesson to fill out the form. In this training, teachers could have hands-on practices on

forehanded (more than one teacher in the classroom) teaching to increase practicality and the ratio of teacher-student contact.

All interviewed observers and the teacher-researcher agreed that DA/TBLI instruction made a difference. Observers noticed that students' involvement and engagement was reflected in their enthusiasm and their heightened classroom energy. One possible explanation for their enthusiasm was their realization that their personal needs were addressed and that they were improving on the ILR scale. They had known that their performance of the formal OPI would be evaluated by the ILR scale. As practical adult learners, this understanding likely elevated their intrinsic motivation to participate in this study and to engage the DA process in the daily lessons. This intrinsic motivation was maximized also by using their critical thinking skills and their own knowledge of the world to speak Arabic for realistic purposes (Brown, 2009; Dean, 2004; Galbraith, 2004a; Long, 2004).

Question 4

The fourth question of this study was: What are the student perceptions of the DA process? The results of the ten 5-point scales and the students' interviews reflected four main themes that would be discussed in this chapter. These four themes were (a) the ability of the DA/TBLI instruction's to diagnose students, (b) DA/TBLI instruction made a difference in the students' learning/performance of Arabic speaking, (c) the ability of the cumulative and concurrent techniques to promote learning, and (d) selecting the input material would be crucial for the DA/TBLI instruction.

Their perception of the DA/TBLI instruction as capable of diagnosing their immature (undeveloped or incomplete) abilities daily and with a high level of accuracy

was due, maybe, to its transparency. The transparency of the hinting system in identifying the level of immaturity for a certain language feature could have been the reason of the students' positive perception. This transparency was made possible by the presentation they attended during the pre-DA phase. In this presentation, the teacher-researcher explained to students the theoretical framework of dynamic assessment. They understood by the end of this presentation that the hints would be standardized and graduate in explicitness. They understood that needing fewer number of hints meant being closer to performing the language feature independently as it was described on the Dynamic Assessment Rubrics Form. They understood the structure of this form and how the teacher-researcher would use it. This understanding may have helped lowering their anxiety, and the teacher-researcher selectiveness of which feature to handle through the dynamic assessment's scaffoldings may have helped lower the affective filter in the classroom (Krashen, 1981). This understanding could have also helped students to diagnose their own needs precisely and to notice their own improvement as the lessons progressed. This self-diagnosing ability was likely another reason for their positive impression about the diagnostic ability of the DA/TBLI approach.

Integrating all language modes in the same lesson daily immersed students in interesting materials and tasks, and this deep involvement consequently might have led to their next perception on which 100% of them agreed: DA/TBLI instruction made a difference. Students expressed greater involvement in lessons of this approach than in their regular-program lessons. They indicated that using interesting material to generate cognitively demanding tasks allowed the process of scaffolding successfully. This teacher-researcher believed that the deep internalization caused by this co-constructivist

method (Poehner, 2005) enabled students to improve in other modes such as listening or reading. The task-based activities prompted students to use input material that was suitable for their current proficiency level in reading, interesting to them, and relevant to the measurable language they needed to generate collaboratively by the end of the task. Repairing errors or helping students overcome their difficulties while being immersed in such a multimodal setting was not only welcomed by students but also was conducive to their advancement in listening and reading.

The scaffolding helped students to improve in more than one mode quickly, because the students were immersed in these scaffoldings in the classroom's concurrent and cumulative techniques of dynamic assessment (Hill & Sabet, 2009). All students agreed in the survey and in their interviews on benefitting from participating directly or indirectly in the scaffolding process with the teacher-researcher. A student expressed their benefitting from scaffolding further by saying that half of the errors needed in the hinting process were common for all students. This remark confirmed the point mentioned earlier about the selectiveness of the teacher-researcher for the features to repair by deploying the hinting process. The teacher of the DA/TBLI instruction would need to select errors that were common and systematic for the whole group to repair through the hinting process. The teacher would need to prioritize error correction by starting with the common errors in the group first before addressing those that would be less common. The performance of features repaired by the scaffolding of dynamic assessment should be improved to the required standards for the targeted sublevel on the Dynamic Assessment Rubrics Form. Engaging students on each one of these errors one by one consecutively soon would lead to the students' independent performance.

Students' advancement in the classroom setting through the concurrent and cumulative techniques of dynamic assessment was enhanced by the multimodal use of Arabic to solve a real-life situation. Progress was furthered by selecting material that was of interest to students per their responses in the biographical questionnaire (Appendix B). The teacher-researcher believed that the material selected would need to be appropriate. Appropriateness in this context referred to the suitability of the input material to the students' present proficiency level and to the assigned outcome of the task used in class. This task outcome that they would need to produce prompted them to collaborate and then present it in Arabic. The teacher-researcher believed that Krashen's (1981) input hypothesis of "i+1" was pivotal in the context of designing a DA/TBLI lesson. Students stated in the interviews that passages that were too difficult needed too long time to process, and consequently the remaining time for speaking and for the hinting process became too short. Therefore, selecting the difficulty level of the input passages based on the principles of text typology is crucial for this process. Students' present listening and reading ability on the ILR scale was the guiding factor to the difficulty level of the input material selected.

The teacher-researcher of the DA/TBLI lessons had to decide whether to use material with a difficulty level matching the students' present proficiency level or to follow Krashen's formula of "i+1." The goal of using the input material was to provide information to prompt students' collaboration for the purpose of generating Arabic through speaking and writing. Therefore, matching their reading and listening proficiency with the used passage's difficulty level would be recommended. This way the time for the lesson would not be consumed for learning the new features in the input material.

Instead, it would be for prompting the hinting process during their collaboration or presenting their final outcome. If the objective, however, was advancing their reading or listening skill as well, then Krashen's "i+1" would be more suitable. In this case, the block of instruction would need to be longer so that students end up with sufficient time to collaborate mainly by speaking in Arabic. Students suggested increasing the time allocated for this approach in their daily lessons.

The input material's topic needed to be known and interesting as expressed clearly by students during their interviews and in their responses to the survey. Combining interesting topics with interesting tasks (Appendix E) enhanced the students' engagement with the material and consequently their collaboration. Being immersed in the task and having fun collaborating to produce the assigned outcome could mitigate their feeling of being on the spot during the scaffolding process. The teacher-researcher found the information gathered from the students' responses to the biographical questionnaire (Appendix B) very helpful in selecting the input material. Students were encouraged also to suggest to the teacher-researcher topics that would be of interest to them at any time during the DA phase. The teacher-researcher communicated to them also that he would welcome feedback from them during the DA phase. He asked students periodically in one-on-one settings what could be improved in the subsequent lessons. This approach enabled the teacher-researcher to select successfully passages in addition to referring to the information gathered from their biographical questionnaires (Appendix B). This open line of communication with students during the DA phase accomplished the purpose of attracting the students to collaborate purposefully. It also elevated their

motivation feeling that they were a part of the planning process (Brown, 2009; Dean, 2004; Galbraith, 2004a; Long, 2004).

Conclusion

This section recaps both types of dynamic assessment and would summarize the findings of this study. It would include the teacher-researcher's suggestions at the end of each idea presented; these ideas would be furthered into possible future researches in the next section. The results of this study would suggest that DA/TBLI instruction would be a successful application of dynamic assessment in a classroom setting. Students' structural control of Arabic improved through the DA phase, and it was reasonable to assume that this improvement was due to the DA/TBLI instruction. In other words, DA/TBLI instruction could promote learning and was capable of diagnosing students' needs in classrooms. The DA type used in the classroom lessons during the DA phase was the DA-interactionist technique (Sternberg & Grigorenko, 2002). This study showed its compatibility with task-based language instruction and its effectiveness for adult learners of Arabic in a classroom setting. DA/TBLI instruction addressed the diversity of students' proficiency level and intellectual styles in the same classroom. It could be effective in adult learners' language classrooms in particular due to its practicality and relevancy to real-life needs.

The other type of DA, the interventionist DA, was used before and after the DA phase to diagnose the language features needing improvements for learners to advance to the higher sublevel in the Dynamic Assessment Form. Interventionist DA (OPI with DA assistance) and the Interactionist DA were much more accurate in diagnosing students' weaknesses and potential learning in accurate details than OPI. The interventionist DA

(OPI with DA assistance), however, had the potential to evaluate Arabic speakers' proficiency level on the ILR scale as an alternative to the OPI instrument. A parallel coefficient of 1.0 was found between pre- and post- OPIs and the DA-interventionist interviews (OPI with DA assistance) in this study. The interventionist interviews still could not be a replacement of the OPI as a summative psychometric test to evaluate the Arabic speaking abilities by the ILR scale. Although it enjoyed high face validity because of the Dynamic Assessment Rubrics Form, it would need a process of maturity and validation to replace a static, valid, reliable, and practical test such as the OPI. As the OPI matured over the years, the same process would be needed for the DA-interventionist interviews to reach the same level of validity and reliability. During this maturation period, an investigation for the parallel statistical coefficient between the regular OPIs and the DA-interventionist interviews for a sufficient number of participants would be crucial.

The technique of establishing the ceiling for the DA-interventionist interviews to evaluate an examinee's proficiency level by the ILR would need further investigation. This study used the assistance provided through the hinting process as the ceiling for the examinee's abilities. This research also would suggest following the exact structure of the OPI with the exception of using the hinting process. The investment of investigating whether the DA-interventionist (OPI with DA assistance) could be used as a summative test is important, because if true, using DA would make language programs more successful and cost effective, that is, this confirmation would mean that dynamic assessment could be used as a summative test, diagnostic test, and in addition to being

effective in classroom settings when combined with task-based language instruction as a successful approach for improving Arabic speaking.

The DA-interventionist, however, had the ability of accurately diagnosing students' needs by using the Dynamic Assessment Rubrics Form, that is, the DA in general and the DA-interventionist in particular had the ability to measure the students' potential learning accurately on the Dynamic Assessment Rubrics Form. The other technique of dynamic assessment known as the interactionist was the one used in this study by combining it with task-based language instruction (TBLI). Combining the DA-interactionist with TBLI was not bound by the OPI tasks as it was the case in the DA-interventionist of this current study. Rather it was based on simulating real-life situations and scenarios to prompt students to use their Arabic authentically and realistically. The results of this study showed DA/TBLI instruction capable of not only diagnosing students' needs but also of promoting learning through improving students' structural control of Arabic.

These results of using DA/TBLI instruction in the Defense Language Institute Foreign Language Center's (DLIFLC) classroom were encouraging and would suggest the efficacy of its use on a wider scale in the Arabic program or for all language programs in DLIFLC. In this case, both students and teachers would need to go through training on its process. Students would need to understand the theoretical framework of dynamic assessment and its hinting process. Teachers would need to have experiential training on dynamic assessment, using the Dynamic Assessment Rubrics Form in class, and on designing DA/TBLI instruction before implementing this technique in their classrooms. The training on designing DA/TBLI lessons should include the selection of

appropriate and interesting input material. In this training, teachers also would need to explore their preferred technique of filling out the Dynamic Assessment Rubrics Form while teaching a lesson and how to select the language features for the DA's hinting process.

Recommendations

This section provides recommendations for DA practices and then provides suggestions for future research. Based on the results of this study, it is the teacher-researcher's belief that the DA/TBLI approach should be implemented for the Arabic Basic Course at the Defense Language Institute in particular and should also be considered for use in select adult language classrooms.

Practices

The interventionist technique should be used first to diagnose the needs and the potential learning for every student. This step would be necessary to guide the placement of students with others who would share the same needs and whose intellectual styles would be compatible as much as possible. The subsequent planning of the course's lessons should consider the students' proficiency levels and needs related to their mature and immature abilities.

These lessons should be designed by combining the interactionist DA with the principles of task-based language instruction as was done in this study. For this purpose, teachers should be trained experientially first on several relevant topics and skills to ensure the success of their teaching efforts. They would need to have intimate understanding of the ILR scale in general and the OPI structure in particular. Then, this intimate understanding of the ILR scale could be transferred to their training on the

Dynamic Assessment Rubrics Form. Developing an effective technique of entering hints into the form while teaching the lesson was important for implementing DA/TBLI instruction successfully.

DA/TBLI instruction, unlike the DA-interventionist interviews, should not be restricted by the OPI prescribed tasks for the different proficiency levels. Designing these lessons, however, would be tedious and labor intensive, because the teacher would need to find suitable material and tasks. The task as a whole would need to be realistic and would include authentic material for interesting topics and at the students' present proficiency level.

To make this process easier for teachers, a repository of material, graphics, multimedia, and their lesson plans could be sorted by lesson in a net-worked learning management system such as Sakai[®] or Blackboard[®]. Applications such as Sakai[®] and Blackboard[®] would be Internet-based and accessible from any geographical location. Using these programs would not only make the material readily accessible to teachers and students but also would enhance students' collaborations synchronously and asynchronously. Students could track their own progress on the Dynamic Assessment Rubrics Form to know which features they would need to improve, and teachers could refer to the same form for their subsequent lesson planning.

Future Research

This study could prompt six future studies. One, the generalizability of this study's findings would need quantitative studies conducted in the future with a sufficient number of participants. This future study should investigate the presence of a statistically significant difference between students' results at the end of the classroom-teaching

phase during which the experimental group, unlike the control group, uses DA/TBLI instruction. This research should be done in a pretest-posttest format in which participants would be evaluated at the beginning and the end by a DA interventionist and OPI interview.

Between these two sets of evaluation, students in two groups of 30 or more should be formed randomly, and divided into control and experiment groups. Both groups would go through the teaching phase in which only the experimental group would be given daily DA/TBLI instruction. This teaching phase in between should be for a sufficient period of time such as the entire third semester (12-16 weeks) at the Defense Language Institute Foreign Language Center. Comparing the means for the two sets of scores for both groups might enable the researcher to find a statistical significant difference between the two means.

Two, a study needed for the future would be for investigating the reliability and the validity of the DA-interventionist interview as a possible alternative or replacement to the OPI instrument. An investigation for the most effective technique of establishing the ceiling of the examinee's abilities could be done by finding the parallel coefficient between a sufficient number of OPIs and interventionist interviews. This process could be done for both techniques of establishing the ceiling. The first would be by considering the hinting process as the ceiling and the other would be continuing the OPI technique of using probes (tasks from the higher proficiency level). The results of the DA-interventionist interviews could be in the form of the examinee's proficiency level on the ILR scale in addition to a table or a narrative about the learners' weaknesses and

strengths. These weaknesses could be expressed by the assisted language features and the number of hints provided for the examinee.

Three, a research study could also investigate if the developed Dynamic Assessment Rubrics Form or a similar instrument would be helpful for students in elementary and middle schools. This suggested study would complement the study done by Lantolf and Poehner (2011) that was reviewed in the second chapter of this current study. The design for this study could be done by using the pretest-posttest format with a sufficient number of elementary school students selected randomly. If the sufficient number of participants is not available, the researcher could use a mixed-method or a qualitative study format.

Four, in addition to Arabic, other studies could be done in the future to investigate if the deconstruction of the ILR would be successful for other languages as well. The study for every language could follow the same design of this current study or a quantitative study as explained previously in the first suggested study, depending on the availability of a sufficient number of students and certified testers.

Five, the deconstruction of other scales such as the guidelines of the American Council on the Teaching of Foreign Languages (ACTFL) could be examined for the DA process. The design for this study could be based on one of the previously suggested settings in this section.

Six, the process of two teachers or more teaching the same lesson (forehanded teaching) would need further investigation for the DA process to explore if students would have their affective filter (anxiety) raised by the process of scaffolding (Krashen, 1981). This study could be a qualitative study or in the format of participatory research.

Limitations of Study

There were three limitations to this study. The first limitation was having six participants only from the Defense Language Institute Foreign Language Center. Although this is the average class size at the Defense Language Institute Foreign Language Center, the situation is different elsewhere. Answering the questions of this study might add to the knowledge accumulated from the previous studies on dynamic assessment. Then eventually, dynamic assessment might be used more systematically on a wider scale at Defense Language Institute Foreign Language Center in particular and in adult language learning classrooms in general. If the findings of this study would show positive results for dynamic assessment in Arabic classrooms, further quantitative studies could be undertaken credibly on a sufficient number of participants for the purpose of generalizing the findings. The inability of generalizing the results of this study was not the only limitation of this study.

The second limitation was that the Arabic variety used in this study was limited to Modern Standard Arabic. Modern Standard Arabic is currently the variety of Arabic mainly taught at the Defense Language Institute Foreign Language Center, and the Arabic dialects taught are still not developed fully for the purpose of this study. Moreover, the testers at the Defense Language Institute Foreign Language Center, 24 of which participated in this study, had been trained for many years on evaluating Modern Standard Arabic systematically. This extensive experience would raise the reliability and the validity of the Oral Proficiency Interview in Modern Standard Arabic (MSA) much more than conducting it for other major dialects. This limitation might be addressed eventually in the future, because Modern Standard Arabic is not the commonly used

variety of Arabic in its speaking countries for most daily tasks; MSA is rather limited to the academic and media purposes.

The third limitation of this study is that the teacher-researcher had been working with the Defense Language Institute Foreign Language Center for 22 years, and his opinions would be influenced by personal views and understandings of the environment. On a positive note, he knew the program intimately after working for such a long time with the Defense Language Institute Foreign Language Center. He would be able to supplement the material to meet the standards discussed in this study so that the results would accurately represent the combining of all the variables mentioned above with dynamic assessment. Knowing the capabilities of the institute's Arabic program would assist in conducting this research to the maximum benefits to the field of Foreign Language Education.

REFERENCES

- Ableeva, R., & Lantolf, J. (2011). Mediated dialogue and the microgenesis of second language listening comprehension. *Assessment in Education: Principles, Policy & Practice*, 18, 133-149.
- American Council on the Teaching of Foreign Languages. (2012). ACTFL Proficiency Guidelines - Speaking (Revised 1999-PDF). Retrieved from <http://www.actfl.org/i4a/pages/index.cfm?pageid=3325>
- Alderson, J. C. (2005). *Diagnosing foreign language proficiency : The interface between learning and assessment*. London, UK: Continuum International Publishing.
- Allal, L., & Pelgrims Ducrey, G. (2000). Assessment "of"- or "in"- the zone of proximal development. *Learning and Instruction*, 10, 137-152.
- Angelo, T., & Cross, K. (1993). *Classroom assessment techniques: A handbook for college teachers* (2nd ed.). San Francisco, CA: Jossey-Bass.
- Anton, M. (2009). Dynamic assessment of advanced second language learners. *Foreign Language Annals*, 42, 576-598.
- Bachman, L. (1990). *Fundamental consideration in language testing*. Oxford, UK: Oxford University Press.
- Bachman, L. (2002). Some reflections on task-based language performance assessment. *Language Testing*, 19, 453-476.
- Bachman, L., & Palmer, A. (1996). *Language testing in practice*. New York, NY: Oxford University Press.
- Baron, M. A., & Boschee, F. (1995). *Authentic assessment: The key to unlocking student success*. Lancaster, PA: Technomic Publishing Co., Inc.
- Bialystok, E. H., K. (1994). *The science and psychology of second-language acquisition*. New York, NY: Basic Books, A Division of Harper Collins Publisher, Inc.

- Bienkowski, S. (2013, 03/22/2013). *Preliminary study on using ILR can do Statements for placement in SOFTS virtual training courses*. Retrieved from http://www.govtilr.org/TC/Presentations/March%202013_Briefs/ILR%20Can%20Do%20Statements%20for%20Placement%20in%20Virtual%20Training%20Courses%20Final%20ILR%20Testing%20Committee%20March%202013.pdf
- Briggs, K. C., & Myers, I. B. (1998). *MBTI self-scorable: Form M*. Menlo Park, CA: CPP, Inc.
- Brown, H. D., & Abeywickrama, P. (2010). *Language assessment: Principles and classroom practices*. New York, NY: Pearson Education, Inc.
- Brown, N. A. (2009). Argumentation and debate in foreign language instruction: A case for the traditional classroom facilitating advanced-level language uptake. *Modern Language Journal*, 93, 534-549.
- Budoff, M. (1987a). Measures for assessing learning potential. In C. S. Lidz (Ed.), *Dynamic testing* (pp. 173-195). New York, NY: Guilford Press.
- Budoff, M. (1987b). The validity of learning potential. In C. S. Lidz (Ed.), *Dynamic testing* (pp. 52-81). New York, NY: Guilford Press.
- Canale, M., & Swain, M. (1980). Theoretical bases of communicative approaches to second language teaching and testing. *Applied Linguistics*, 1(1), 1-47.
- Carlson, J. S., & Wiedel, K. H. (1978). Use of testing-the-limits procedures in the testing of intellectual capabilities in children with learning difficulties. *American Journal of Mental Deficiency*, 11, 559-564.
- Child, J. R. (1987). Language proficiency and the typology of texts. In H. Byrnes & M. Canale (Eds.), *Defining and developing proficiency: Guidelines, implementation, and concept*. Lincolnwood, IL: National Textbook Co.
- Child, J. R. (1998). Language skill levels, textual modes, and the rating process. *Foreign Language Annals*, 31, 381-391.
- Child, J. R. (2001). Analysis of texts and critique of judgment. In J. E. Alatis & A. H. Tan (Eds.), *Georgetown University Round Table on Languages and Linguistics 1999: Language in Our Time: Bilingual Education and Official English, Ebonics*

and Standard English, Immigration and the Unz Initiative Languages and Linguistics 1999 (pp. 88-98). Washington, DC: Georgetown University Press.

Child, J. R., Clifford, R., & Lowe, P. (1993). Proficiency and performance in language testing. *Applied Language Learning*, 1(4), 19-54.

Child, J. R., & Lowe, P. (1998). *Density & syntax*. Workshop material. National Cryptological School. Washington, DC.

Chomsky, N. (1968). *Language and mind*: New York, NY, Harcourt.

Creswell, J. W. (2007). *Qualitative inquiry and reseach design: Choosing among five approaches* (2nd ed.). Thousand Oak, CA: Sage Publications, Inc.

Dean, G. J. (2004). Designing instruction. In M. W. Galbraith (Ed.), *Adult learning methods* (3rd ed., pp. 93-118). Malabar, FL: Krieger Publishing Company.

Dewey, J. (1963). *Experience and education*: New York, NY: Collier Books.

Defense Language Institute Foreign Language Center. (2010). *OPI 2000: Tester certification workshop training manual*. Monterey, CA: Proficiency Standards Division

Defense Language Institute Foreign Language Center. (2013a). *Home*. Retrieved from <http://www.dliflc.edu/index.html>

Defense Language Institute Foreign Language Center. (2013b). *Test Management Division*. Retrieved from <http://www.dliflc.edu/testmanagementdi.html>

Defense Language Institute Foreign Language Center. (2013c). *About DLIFLC*. Retrieved from <http://www.dliflc.edu/about.html>

Defense Language Institute Foreign Language Center. (2013d). *Proficiency standards division*. Retrieved from <http://www.dliflc.edu/proficiencystand.html>

Doolittle, P. E. (1995, June). *Understanding cooperative learning through Vygotsky's zone of proximal development*. Paper presented at the Lilly National Conference on Excellence in College Teaching, Columbia, SC. <http://0->

search.ebscohost.com.ignacio.usfca.edu/login.aspx?direct=true&db=eric&AN=E
D384575&site=ehost-live&scope=site

- Doolittle, P. E. (1997). Vygotsky's zone of proximal development as a theoretical foundation for cooperative learning. *Journal on Excellence in College Teaching*, 8(1), 83-103.
- Dunn, W. E., & Lantolf, J. P. (1998). Vygotsky's zone of proximal development and Krashen's "i + 1": Incommensurable constructs; incommensurable theories. *Language Learning*, 48, 411-442.
- Edwards, D. (2004). The role of languages in post-9/11 United States. *Modern Language Journal*, 88, 268-271.
- Ellis, R. (2005). Planning and task-based research: Theory and research. In R. Ellis (Ed.), *Planning and task performance in a second language* (pp. 3-34). Philadelphia, PA: John Benjamins Publishing Company.
- Ellis, R. (2009a). The differential effects of three types of task planning on the fluency, complexity, and accuracy in L2 oral production. *Applied Linguistics*, 30, 474-509.
- Ellis, R. (2009b). *Task-based language learning and teaching* (6th ed.). Oxford, UK: Oxford University Press.
- Fan, W., & He, Y. (2012). Academic achievement and intellectual styles. In L. F. Zhang, R. J. Sternberg, & S. Rayner (Eds.), *Handbook of intellectual styles: Preferences in cognition, learning, and thinking* (pp. 233-249). New York, NY: Springer Publishing Company.
- Ferrara, R. A., Brown, A. L., & Campione, J. C. (1986). Children's learning and transfer of inductive reasoning rules: Studies in proximal development. *Child Development*, 57, 1087-1099.
- Foster, P., & Skehan, P. (1996). The influence of planning and task type on second language performance. *Studies in Second Language Acquisition*, 18, 299-323.
- Foster, P., & Skehan, P. (1999). The influence of source of planning and focus of planning on task-based performance. *Language Teaching Research*, 3, 215-247.

- Foreign Service Institute. (2013). *Foreign Service Institute*. Retrieved 07/14/2013, 2013, from <http://www.state.gov/m/fsi/>
- Furman, N., Goldberg, D., & Lustin, N. (2010). Enrollments in languages other than English in United States institutions of higher education, Fall 2009 Retrieved 07/14/2013, 2013, from http://www.mla.org/pdf/2009_enrollment_survey.pdf
- Galbraith, M. W. (2004a). *Adult learning methods: A guide for effective instruction*. Malabar, FL: Krieger Publishing Company.
- Galbraith, M. W. (2004b). The teacher of adults. In M. W. Galbraith (Ed.), *Adult learning methods: A guide for effective instruction* (3rd ed., pp. 3-21). Malabar, FL: Krieger Publishing Company.
- Gnadinger, C. M. (2008). Peer-mediated instruction: Assisted performance in the primary classroom. *Teachers and Teaching: Theory and Practice*, 14, 129-142.
- Goos, M., Galbraith, P., & Renshaw, P. (2002). Socially mediated metacognition: Creating collaborative zones of proximal development in small group problem solving. *Educational Studies in Mathematics*, 49, 193-223.
- Grigorenko, E. L., & Sternberg, R. J. (2002). *Dynamic testing: The nature and measurement of learning potential*. Cambridge, UK: Cambridge University Bridge.
- Guthke, J., & Stein, H. (1996). Are learning tests the better version of intelligence tests? *European Journal of Psychological Assessment*, 12, 1.
- Havnes, A. (2008). Peer-Mediated Learning beyond the Curriculum. *Studies in Higher Education*, 33, 193-204.
- Hill, K., & Sabet, M. (2009). Dynamic speaking assessments. *TESOL Quarterly: A Journal for Teachers of English to Speakers of Other Languages and of Standard English as a Second Dialect*, 43, 537-545.
- Interagency Language Roundtable. (2013a). *Interagency language roundtable language skill level descriptions - speaking*. Retrieved from <http://www.govtilr.org/Skills/ILRscale2.htm>

- Interagency Language Roundtable. (2013b). *An overview of the history of the ILR Language proficiency skill level descriptions and scale* by Dr. Martha Herzog Retrieved from <http://www.govtilr.org/Skills/IRL%20Scale%20History.htm>
- Johnson, L. R., Penny, A. J., & Gordon, B. (2009). *Assessing performance: Designing, scoring, and validating performance tasks*. New York, NY: The Guilford Press.
- Kinging, C. (2002). Defining the zone of proximal development in US foreign language education. *Applied Linguistics*, 23, 240-261.
- Krashen, S. D. (1982). *Principles and practice in second language acquisition*. Oxford, UK: Pergamon.
- Lantolf, J. P. (2009). Dynamic assessment: The dialectic integration of instruction and assessment. *Language Teaching*, 42, 355-368.
- Lantolf, J. P., & Poehner, M. E. (2007). *Dynamic assessment in the foreign language classroom: A teacher's guide*. University Park, PA: CALPER Publications.
- Lantolf, J. P., & Poehner, M. E. (2009). Dynamic Assessment in the Classroom: Vygotskian Praxis for Second Language Development. *Language Teaching Research*, 15(1), 11-33.
- Lantolf, J. P., & Poehner, M. E. (2011). Dynamic assessment in the classroom: Vygotskian praxis for second language development. *Language Teaching Research*, 15(1), 11-33.
- Lantolf, J. (2012, 03/29/2012). *The pedagogical imperative and the dialectics of L2 instruction*. Paper presented at the TESOL International Convention & English Language Expo, Philadelphia, PA.
- Larsen-Freeman, D. (1991b). Second language acquisition: Staking out the territory. *TESOL Quarterly: A Journal for Teachers of English to Speakers of Other Languages and of Standard English as a Second Dialect*, 25, 315-350.
- Lidz, C. S. (1991). *Practitioner's guide to dynamic assessment*. New York, NY: Guilford.

- Long, H. B. (2004). Understanding adult learners. In M. W. Galbraith (Ed.), *Adult learning methods: A guide for effective instruction* (pp. 23-37). Malabar, FL: Krieger Publishing Company.
- Long, M. H. (2000). Focus on form in task-based language teaching. In R. D. Lambert & E. Shohamy (Eds.), *Language Policy and Pedagogy* (pp. 179-192). Philadelphia, PA: Benjamins.
- Lowe, P. (2000). *James Child's text modes & their derivatives: A compilation of descriptions*. Washington, DC: National Cryptological School.
- Messick, S. (1989). Validity and washback in language testing. *Language Testing*, 13, 241-256.
- Messick, S. (1994). The interplay of evidence and consequences in the validation of performance assessment. *Educational Researcher*, 23, 13-23.
- Nunan, D. (2004). *Task-based language teaching*. Cambridge, UK: Cambridge University Press.
- Piaget, J. (1971). *Biology and knowledge: An essay on the relations between organic regulations and cognitive processes*. Chicago, IL: The University of Chicago Press.
- Poehner, M. E. (2005). *Dynamic assessment of oral proficiency among advanced L2 learners of French*. (Ph.D. 3193226), The Pennsylvania State University, United States -- Pennsylvania. Retrieved from <http://0proquest.umi.com.ignacio.usfca.edu/pqdweb?did=1008320051&Fmt=7&clientId=16131&RQT=309&VName=PQD>
- Poehner, M. E. (2009). Group dynamic assessment: Mediation for the L2 classroom. *TESOL Quarterly: A Journal for Teachers of English to Speakers of Other Languages and of Standard English as a Second Dialect*, 43, 471-490.
- Poehner, M. E., & Lantolf, J. P. (2005). Dynamic assessment in the language classroom. *Language Teaching Research*, 9, 233-265.

- Poehner, M. E., & Lantolf, J. P. (2010). Vygotsky's teaching-assessment dialectic and L2 education: The case for dynamic assessment. *Mind, Culture, and Activity*, 17, 312-330.
- Skehan, P. (1996). A framework for the implementation of task-based instruction. *Applied Linguistics*, 17(1), 38-62.
- Skehan, P. (1998). Task-based instruction. *Annual Review of Applied Linguistics*, 18, 268-286.
- Skehan, P., & Foster, P. (1999). The influence of task structure and processing conditions on narrative retellings. *Language Learning*, 49, 93-120.
- Skorton, D., & Altschuler, G. (2012). America's foreign language deficit Retrieved 07/14/2013, 2013, from <http://www.forbes.com/sites/collegeprose/2012/08/27/americas-foreign-language-deficit/>
- Sternberg, R. J., Wagner, R. K., & Zhang, L. F. (2007). Thinking styles inventory - Revised II *Thinking styles inventory - Revised II B2 - Thinking styles inventory - Revised II*: Tufts University.
- Stryker, S., & Leaver, B. (1997). Content-based instruction: From theory to practice. In S. Stryker & B. Leaver (Eds.), *Content-based instruction in foreign language education. Models and methods* (pp. 3-28). Washington, DC: Georgetown University Press.
- Swain, M., Kinnear, P., & Steinman, L. (2010). *Sociocultural theory in second language education: An introduction through narratives*: Multilingual Matters.
- van Lier, L. (1996). *Interaction in the language curriculum: Awareness, autonomy & authenticity*. New York, NY: Longman Group Limited.
- Vygotsky, L. (1978). *Mind in society: The development of higher psychological process*. Cambridge, MA: Harvard University Press.
- Wesche, M. (2004). Teaching languages and cultures in post-9/11 world. *Modern Language Journal*, 88, 275-285.

- Zhang, L. F., & Sternberg, R. J. (2005). A threefold model of intellectual styles. *Educational Psychology Review*, 17, 1-53. doi: 10.1007/s10648-1635-4
- Zhang, L. F., Sternberg, R. J., & Rayner, S. (2012). Intellectual styles: challenges, milestones, and agenda. In L. F. Zhang, R. J. Sternberg, & S. Rayner (Eds.), *Handbook of intellectual styles: Preferences in cognition, learning, and thinking* (pp. 1-20). New York, NY: Springer Publishing Company.

APPENDIX A
INFORMED CONSENT FORM
UNIVERSITY OF SAN FRANCISCO

INFORMED CONSENT FORM

UNIVERSITY OF SAN FRANCISCO

CONSENT TO BE A RESEARCH SUBJECT - STUDENT

Purpose and Background

Xxxxx X Xxxxx, a doctoral student in the School of Education at the University of San Francisco is conducting a study on Arabic adult learners attending the Defense Language Institute. The researcher will explore the impact of using Dynamic Assessment (DA) in a classroom setting on the speaking progress of students in Semester III of the Arabic Basic Course.

I am being asked to participate because I meet the following criteria:

- (a) I am a student in DLIFLC
- (b) I am a student in the Arabic Basic Course
- (c) I am a student in Semester III
- (d) I have already been through ICPT 301

Procedures

If I agree to be a part of this study, the following will happen:

1. I will attend a presentation about Dynamic Assessment and how its rubrics will be used in class.
2. I will answer questionnaires about my intellectual styles (learning styles and personality traits), sensory preference, and background information.
3. I will attend one-hour Arabic lesson starting at the beginning of 302 until graduation.
4. The researcher will be the teacher of the Arabic lesson mentioned above in item # 3.
5. During my Arabic lessons, I will do peer-assessment as explained to me in the presentation mentioned above.
6. I will participate in an OPI prior to the beginning of classes
7. I will receive a DA prior, during, and at the end of classes and prior to graduation.
8. I will respond to a questionnaire at the end of this study soliciting my opinion about DA.

Risks/Discomforts

1. During these Arabic lessons, I will be prompted to represent my small working group to the rest of the six-student class, but I can always decline playing this part.
2. Sometimes the critical thinking required in the daily activities will prompt me to share my opinion about issues that might make me uncomfortable, but I can

always decline voicing my real opinion at that time whether to my small working group or to the whole class.

3. I might feel uncomfortable conducting peer-assessment activities, but I can always stop allowing my work to be evaluated by a peer.
4. I might not feel comfortable with integrating more than one skill (Listening, Reading, Speaking, and Writing) in every lesson of this study.
5. I might not feel comfortable with an observer coming to class once or twice weekly; these are DLIFLC certified testers who are giving the researcher feedback and they will be invisible in the classroom

Benefits

The direct benefit to you is having the opportunity to practice through an OPI, get accurate diagnosis to your progress on the ILR, and to learn Arabic through well prepared and tailored lessons. These lessons are targeting the improvement of your speaking ability by being designed according to the latest in the field of Second Language Acquisition.

The anticipated benefit of this study is informing DLIFLC and the field of foreign language teaching with the findings of this study. These findings will eventually prompt others to conduct the quantitative study required for generalizing the results. Then, many adult learners of Arabic in particular and a foreign language at large will benefit from your participation in this study.

Alternative

I am free not to participate in this study.

Costs/Financial Considerations

There will be no financial costs to be charged for my participation in this study.

Reimbursement

I will not be reimbursed or paid for my participation in this study.

Questions

I have talked with Xxxxx X. Xxxxx about this study, and have had my questions answered. If I have any further questions about the study, I may call him on his cell phone (xxx) xxx-xxxx or email him at xxxxxxxxxxxxxxxx@gmail.com.

If I have any questions or comments about participating in this study, I should first talk to the researcher. If for some reason I don't wish to do this, I may contact the IRBPHS, which is concerned with the protection of volunteers in research projects. I may reach the IRBPHS office by calling xxx-xxx-xxxx and leaving a voice mail message, by e-mailing XXXXXX@usfca.edu, or by writing to the IRBPHS, Department of International and Multicultural Education, Education Bldg., University of San Francisco, 2130 Fulton Street, San Francisco, CA 94117-1080.

Consent

I have been given a copy of this signed consent form to keep.

PARTICIPATION IN RESEARCH IS VOLUNTARY. I am free to decline to be in this study or to withdraw from it at any point. My decision as to whether or not to participate in this study will have no influence on my present or future status as a student in DLIFLC or as an American soldier. My signature below indicates that I agree to participate in this study.

Subject Signature

Date of Signature

Person obtaining consent, XXXXX X. XXXXX

Date of Signature

Appendix B
Biographical Questionnaire

Biographical Questionnaire

Please complete all relevant items to provide some background information on yourself and some factors related to your language-learning circumstances:

NAME AND RANK	
SERVICE	
UNIT	
DOB	
MARITAL STATUS	

1. Use this space to tell us about your family (father, mother, siblings, spouse, and maybe your kids). You can mention work, education, or ages for siblings and kids.

--

2. What is your highest level of education? If you attended college, what was your major?

3. What other languages have you learned in addition to English and Arabic?

--

4. How did you learn any other language?

--

**5. What do think your proficiency level is in all the foreign languages you know
(Listening, Reading, Speaking, and Writing)?**

--

6. What are your hobbies and topics of interests?

7. Have you traveled internationally or domestically? Write briefly about those trips.

- 8. If you haven't traveled internationally or domestically, are you interested in traveling and going on trips in the future? Where? To do what?**

- 9. What other employment (or volunteer) had you experienced before joining the military?**

- 10. Would it be interesting to use Arabic material pertinent to your background hobbies, work experiences, and trips?**

11. What are your suggestions on classroom material and activities?

Additional Comments (you can use the back for additional space):

Appendix C

Biographical Questionnaire

Dynamic Assessment Rubrics Form (DARF) for Teachers

Gradual Hints for the ILR Descriptors

Dynamic Assessment Rubrics Form (DARF) for Teachers

Gradual Hints for the ILR Descriptors

Name/s:

Initial Performance					
Level of explicitness					
Performance					
Level of explicitness					
Performance					
Level of explicitness					

Remarks (you can use the back or include in your own observation notes):

Appendix D

Dynamic Assessment Rubrics Form (DARF) for Observers

Gradual Hints for the ILR Descriptors

Dynamic Assessment Rubrics Form (DARF) for Observers

Gradual Hints for the ILR Descriptors

Name/s:

Initial Performance	Accuracy in basic grammatical relations is evident. May exhibit the more common forms of verb tenses, for example, but may make frequent errors in formation and selection. While some structures are established, errors occur in more complex patterns. The individual cannot sustain coherent structures in longer utterances or unfamiliar situations. Person, space, and time references are often used incorrectly.	Shows control of all tenses most of the time. Utterances are minimally cohesive most of the time. Basic grammatical structures are typically controlled most of the time. Person, space, and time references are often used correctly. Can sustain coherent structures in longer utterances.	Shows control of all tenses. Utterances are minimally cohesive. Basic grammatical structures are typically controlled. Person, space, and time references are often used correctly. Can sustain coherent structures in longer utterances.	Shows control of all tenses. Utterances are cohesive most of the time. Basic grammatical structures are typically controlled. Person, space, and time references are often used correctly. Can sustain coherent structures in longer utterances.	Shows control of all tenses. Utterances are cohesive most of the time. Basic grammatical structures are controlled most of the time. Person, space, and time references are used correctly. Can sustain coherent structures in longer utterances.
Level of explicitness					
Performance					
Level of explicitness					
Performance					
Level of explicitness					

Remarks (you can use the back or include in your own observation notes):

Appendix E
LESSON PLANS

LESSON PLAN

Teacher's Name: Mohsen Fahmy	Date: 5/16/2013	Time: 2:00 to 2:50 PM
Week of Instruction: 54	Rm #: 331	

Topic/ Subject Area: The Iranian Nuclear Program

Skills Covered: Reading, Listening, Speaking, and Writing

Learning Objectives: By the end of this hour, students will be able to explain to their counterparts from an Arab country the US position on the Iranian nuclear ambitions.

Major Vocabulary Words:

أفرزت تداعيات ، العدول عن سعيها ، تخصيب اليورانيوم ، أخفقت الجهود الدبلوماسية ، أخفقت الجهود الدبلوماسية ،
التخلي عن ، أقامت الدنيا ولم تقعد ، أصرت إيران على موقفها ، يأخذ هذا الأمر على محمل الجد

Major Grammar Points (if applicable): Reviewing the basic grammatical features of Arabic as described for proficiency Level 2 in the ILR.

Time	Action	Materials Used
7 min	Lead in: Warm-Up/ Brainstorming	http://youtu.be/sUdrEd9yQvo
15 min	<ul style="list-style-type: none"> Students will watch a one-minute video to attract their attention to the day's topic. In small groups of three, students will exchange their existing knowledge about the Iranian nuclear program and the position of the US, Israel, and the surrounding Arab countries in this regard. The two groups present their findings. <p>Presentation:</p> <ul style="list-style-type: none"> In two groups, students read the provided handout to guess the meaning of the underlined new vocabulary items. Each group will use each new vocabulary item or phrase to write a new sentence on the classroom's white boards Students will critically review each other's sentences 	
15 min	Practice:	

13 min	<ul style="list-style-type: none">• Students will listen to the text twice and take notes• In their small groups, students will use their notes to answer the content questions provided on page 31.11 and 31.12 <p>Task:</p> <p>Assuming that while being deployed in one of the friendly Arab countries, you were asked about your take on the Iranian nuclear ambition. Discuss with your group the best possible response. Each group presents their suggested response to the other group while the other group plays the role of the Arab counterpart.</p>	
--------	--	--

LESSON PLAN

Teacher's Name: Mohsen Fahmy	Date: 5/17/2012	Time: 9:55 to 10:45
Week of Instruction: 54	Rm #: 331	

Topic/ Subject Area: The Iranian Nuclear Program

Skills Covered: Reading, Speaking, Writing, and Listening

Learning Objectives: By the end of this hour, students will be able to explain to their counterparts from an Arab country the US position on the Iranian nuclear ambitions.

Major Vocabulary Words:

أفرزت تداعيات ، العدول عن سعيها ، تخصيب اليورانيوم ، أخفقت الجهود الدبلوماسية ، أخفقت الجهود الدبلوماسية ،
التخلي عن ، أقامت الدنيا ولم تقعدھا ، أصرت إيران على موقفھا ، يأخذ هذا الأمر على محمل الجد

Major Grammar Points (if applicable): Reviewing the basic grammatical features of Arabic as described for proficiency Level 2 in the ILR.

Time	Action	Materials Used
5 min	Lead in: Warm-Up/ Brainstorming	
20 min	<ul style="list-style-type: none"> In small groups of three, students will exchange their existing knowledge about the Iranian nuclear program and the position of the US, Israel, and the surrounding Arab countries in this regard. <p>Presentation:</p> <ul style="list-style-type: none"> Students read the handouts to guess the meaning of the underlined words to use in a sentence of their own In a round robin, students tell the class their new sentences Students listen to passages 1A, and read 1B and 2A to synthesize their content on the board in their own word 	
25 min	<p>Task:</p> <p>Assuming that while being deployed in one of the friendly Arab countries, you were asked about your take on the Iranian nuclear ambition. In two groups, students will debate the two possible solutions. Impose the will of the international community on Iran vs. the denuclearization of the Middle East</p>	

LESSON PLAN

Teacher's Name: Mohsen Fahmy	Date: 5/20/2012	Time: 9:55 to 10:45
Week of Instruction: 54	Rm #: 331	

Topic/ Subject Area: The Syrian Current Development

Skills Covered: Reading, Speaking, Writing, and Listening

Learning Objectives: By the end of this hour, students will be able to debate their different stands on the Syrian current events.

Major Vocabulary Words:

Major Grammar Points (if applicable): Reviewing the basic grammatical features of Arabic as described for proficiency Level 2 in the ILR.

Time	Action	Materials Used
5 min	Lead in: Warm-Up/ Brainstorming	
20 min	<ul style="list-style-type: none"> Students in pair talk about the Syrian current events Presentation: <ul style="list-style-type: none"> Each pair reads a different passage on Syria In two groups, students synthesize the content of the three passages into one summary. They write bullets of this summary on the board. Each pair reports their summary to the whole group Task:	
25 min	A debate between the two groups. One group sees to it that the US should arm the Syrian Free Army and impose economic sanctions on Syria. The other group sees to it that negotiation between the different Syrian parties under the auspices of the international community is the best approach	

LESSON PLAN

Teacher's Name: Mohsen Fahmy	Date: 5/21/2012	Time: 9:55 to 10:45
Week of Instruction: 54	Rm #: 331	

Topic/ Subject Area: Tourism

Skills Covered: Reading, Speaking, Writing, and Listening

Learning Objectives: By the end of this hour, students will be able to report on a plan for boosting tourism in Monterey, Carmel, and Pacific Grove.

Major Vocabulary Words:

Major Grammar Points (if applicable): Reviewing the basic grammatical features of Arabic as described for proficiency Level 2 in the ILR.

Time	Action	Materials Used
5 min	Lead in: Warm-Up/ Brainstorming	
	<ul style="list-style-type: none"> In pairs, students talk about the annual Academy Award (Oskar), and then report to the whole class what they discussed. 	
25 min	<p>Presentation:</p> <ul style="list-style-type: none"> Each pair reads a different passage on Kan Festival In pairs, develop a summary for their passage by discussing its content. Each pair rehearse how they will report their finding to another group of students In two groups, students synthesize the information from the three passages to report their summary to the whole class. 	
20 min	<p>Task:</p> <p>The cities of Monterey, Carmel, and Pacific Grove are discussing a project of hosting the Annual Academy Award in the Area. One of the suggestions is to realign DLIFLC to elsewhere and use its location for the project. Discuss in your group the best plan possible for using DLIFLC's location and the area at large for hosting the annual event in Monterey for the purpose of boosting its tourism.</p>	

LESSON PLAN

Teacher's Name: Mohsen Fahmy	Date: 5/22/2012	Time: 9:55 to 10:45
Week of Instruction: 54	Rm #: 331	

Topic/ Subject Area: Planning Immersion

Skills Covered: Speaking, Writing

Learning Objectives: By the end of this hour, students will be able to report their plan for improving the immersion program in the target country.

Major Vocabulary Words:

Major Grammar Points (if applicable): Reviewing the basic grammatical features of Arabic as described for proficiency Level 2 in the ILR.

Time	Action	Materials Used
10 min	<u>Lead in: Warm-Up/ Brainstorming</u>	
	<ul style="list-style-type: none"> Students share with each other as much as possible the best or worst immersion experience they had in UMA 	
20 min	<u>Pre-Task1:</u> The teacher will open the door for the issues that they need to discuss time constraints, budget, simulating real life ideas, and ... etc. <u>Task 1:</u> <ul style="list-style-type: none"> In your small group, discuss a plan for improving the current immersion program. Each group writes bullets for their plan on the white board. Then, the group members take turns to present their plan. 	
20 min	<u>Pre-Task 2:</u> The teacher will shed the light on the issues they need to discuss such as safety, the coordination efforts, exploiting their presence in the target country, and defending their selections and the purpose of their activities.	

	<p><u>Task 2:</u></p> <p>You have been tasked to improve the OCONUS immersion. Share your suggestions with your group members to present your plan to the whole class. Use your knowledge and imagination to select a country, the students' activities during AM and PM hours of the immersion days. The two groups will merge their plan into one final plan.</p>	
--	--	--

LESSON PLAN

Teacher's Name: Mohsen Fahmy	Date: 5/28/2012	Time: 9:55 to 10:45
Week of Instruction: 56	Rm #: 331	

Topic/ Subject Area: Planning the construction of a building for a language school

Skills Covered: Speaking, Writing

Learning Objectives: By the end of this hour, students will be able to present their plan for the construction of a language school building.

Major Vocabulary Words:

Major Grammar Points (if applicable): Reviewing the basic grammatical features of Arabic as described for proficiency Level 2 in the ILR.

Time	Action	Materials Used
5 min	<u>Lead in: Warm-Up/ Brainstorming</u>	
15 min	<ul style="list-style-type: none"> Students share with each other as much as possible the description of the best or the worst place they visited or saw in the past <u>Step 1:</u> <ul style="list-style-type: none"> In pairs, students discuss with each other the perfect settings and floor plans of a language-school building. Each pair presents their description of this building. 	
15 min	<u>Step 2:</u> The commandant of DLIFLC asked UMA students to present the Commandant of the Egyptian DLIFLC with a suggestion for the best setting, format, and floor plan of a language school building. This school has five departments each of which has four teams of six teachers. Each team is responsible of teaching English as a second language to 30 Egyptian military students.	
15 min	<u>Step 3:</u> The two groups merge their plans into one final plan and present it to the Egyptian General. The teacher plays the role of the Egyptian General.	

LESSON PLAN

Teacher's Name: Mohsen Fahmy	Date: 5/29/2012	Time: 9:55 to 10:45
Week of Instruction: 56	Rm #: 331	

Topic/ Subject Area: Planning the instructions needed for the risk control management of the language school building designed in the previous lesson

Skills Covered: Speaking, Writing

Learning Objectives: By the end of this hour, students will be able to present their plan for the risk-control-management instructions required for the language school building designed in the previous lesson.

Major Vocabulary Words:

Major Grammar Points (if applicable): Reviewing the basic grammatical features of Arabic as described for proficiency Level 2 in the ILR.

Time	Action	Materials Used
10 min	<u>Lead in: Warm-Up/ Brainstorming</u>	
	<ul style="list-style-type: none"> Students share with each other as much as possible the different steps of risk management for our school building and their barracks. 	
10 min	<u>Step 1:</u>	
	<ul style="list-style-type: none"> In different pairs, students discuss with each other the risk management instructions for the language school building they designed yesterday. 	
15 min	<u>Step 2:</u>	
	<ul style="list-style-type: none"> The commandant of DLIFLC also tasked the same group of students from UMA to create the proper posters, flayers, and documents in Arabic as suggested models for the Egyptian General. Students decided to work in two groups first. 	
15 min	<u>Step 3:</u>	
	<ul style="list-style-type: none"> The two groups merge their plans into one final plan and present it to the Egyptian General. The teacher will play the role of the Egyptian General. 	

LESSON PLAN

Teacher's Name: Mohsen Fahmy	Date: 5/30/2012	Time: 9:55 to 10:45
Week of Instruction: 56	Rm #: 331	

Topic/ Subject Area: The Ethiopian dam on the Blue Nile

Skills Covered: Reading, Speaking, Writing, and Listening

Learning Objectives: By the end of this hour, students will be able to report the latest development of the Ethiopian dam. This report will lead the class to debate how the Egyptian reaction should be.

Major Vocabulary Words:

Major Grammar Points (if applicable): Reviewing the basic grammatical features of Arabic as described for proficiency Level 2 in the ILR.

Time	Action	Materials Used
5 min	Lead in: Warm-Up/ Brainstorming	
	<ul style="list-style-type: none"> In pairs, students talk about their current knowledge about building the dam on the Blue Nile in Ethiopia. Each pair will share their information with the whole class. 	
	Pre-task	
10 min	Their presentation of their current knowledge on the subject will lead to a quick presentation from the teacher on the issue to include the new vocabulary items	
	Presentation:	
20 min	<ul style="list-style-type: none"> Each pair reads a different passage on the topic, and then the students of each pair exchange their understanding of the article to create a summary for it in bullets. In two groups, students synthesize the content of the three passages into one summary. They write bullets of this summary on the board. Each group reports their summary to the other group 	
25 min	Pre-task: The teacher explains the situation of building this dam and its dangerous consequences on the whole area.	

	<p>Production:</p> <p>Task: Should Egypt negotiate a deal with Ethiopia? What if Ethiopia continued regardless of the Egyptian concerns? Discuss the answer with your group, and then present to the whole class with your justification for this opinion. Both groups will defend their stand in a debate.</p>	
--	---	--

Appendix F

Dynamic Assessment Rubrics Form (DARF) for Observers

Gradual Hints for the ILR Descriptors

Verbal
Sentence (vs)
Dual (d)
Plural (plu)
Prepositions
(pre)

Noun in Construct (nc)
Conjugating Present Tense (conj prs)
Conjugating Past Tense (conj pst)
Present Tense after أن (P/أن)

Dynamic Assessment Rubrics Form (DARF) for Observers

Gradual Hints for the ILR Descriptors

Name	Performance	Accuracy in basic grammatical relations is evident. May exhibit the more common forms of verb tenses, for example, but may make frequent errors in formation and selection. While some structures are established, errors occur in more complex patterns. The individual cannot sustain coherent structures in longer utterances or unfamiliar situations. Person, space, and time references are often used incorrectly.	Shows control of all tenses most of the time. Utterances are minimally cohesive most of the time. Also, basic grammatical structures are typically controlled. Person, space, and time references are often used correctly. Can sustain coherent structures in longer utterances.	Shows control of all tenses. Utterances are minimally cohesive. Also, basic grammatical structures are typically controlled. Person, space, and time references are often used correctly. Can sustain coherent structures in longer utterances.	Shows control of all tenses. Utterances are cohesive most of the time. Also, basic grammatical structures are typically controlled. Person, space, and time references are often used correctly. Can sustain coherent structures in longer utterances.	Shows control of all tenses. Utterances are cohesive most of the time. Also, basic grammatical structures are controlled most of the time. Person, space, and time references are used correctly. Can sustain coherent structures in longer utterances.
------	-------------	---	---	---	--	---

Jamal (جمال)	Features and Number of Hints					
Ibrahim (إبراهيم)	Features and Number of Hints					
Hazem (حازم)	Features and Number of Hints					
Ramzy (رمزي)	Features and Number of Hints					

Salwa (سلوى)	Features and Number of Hints					
Basem (باسم)	Features and Number of Hints					