The Effectiveness of Diagnostic Assessment on the Development of Turkish Language Learners' Narrative Skills as an Oral Proficiency Interview (OPI) Task

Anjel Tozcu
Defense Language Institute Foreign Language Center

Abstract

This study investigated the effectiveness of diagnostic assessment on improving students' proficiency in narrating past events, an Oral Proficiency Interview (OPI) Level 2 task. It found that students who were given a personalized learning plan subsequent to the diagnostic assessment interview significantly improved their proficiency in basic sentence structures than those in a control group. They used a significantly larger number of cohesive devices as compared to the control group and exhibited significantly increased accuracy in using cohesive devices than a control group. The students in the treatment group worked on the recommended activities based on the data gathered during the diagnostic assessment interview and the pre-interview questionnaires, i.e., the E & L, MBTI, and Barsch. The students in the control group spent the same amount of time reading narrations, doing comprehension exercise,s and following standard teacher feedback for improvement. Although both groups showed increases in accurate use of cohesive devices and proficiency in basic sentence structures, the treatment students showed significantly greater gains than the control students.

1. Introduction

Language teachers have always known the importance of diagnosing their students' strengths and weaknesses early on in a language course in order to tailor teaching and plan classroom activities to accommodate students (Reed, 2006). However, this is one of those "pedagogical insights" that is easier said than acted upon (p. 1). Many practitioners disregard diagnostic assessment and simply teach their class with the textbook they happen to be using and carry out activities that have in the past worked well for them—believing that they will know their students pretty well as the course progresses. The problem is that while students' strengths and weaknesses may gradually become more obvious toward the end of the course, by the time teachers get to know strengths and weaknesses of their students, it is usually too late to take action (2006). Thus, Pimsleur began to underline the significance of "knowing your students in advance" (Pimsleur & Struth, 1968). The importance of diagnostic assessment in language teaching has recently been highlighted (Alderson, 2006). The Diagnostic Assessment Center of the Defense Language Institute Foreign Language Center (2006), Leaver, Ehrman and Shekhtman (2005), Leaver (1998), and Ehrman (1996) created a list of features characterizing diagnostic approaches in language teaching.

Diagnostic assessment (DA) is a formative evaluation of a learner's their strengths and weaknesses in speaking, reading, and listening comprehension, (Leaver & Campbell, 2015; Defense Language Institute Foreign Language Center Directorate of Continuing Education Educational Support Services Diagnostic Assessment Center, Diagnostic Assessment Workshop, 2007). It thoroughly indicates which specific areas of work a student needs to surpass. Diagnostic assessment within the context of language learning provides learners with meaningful feedback and recommendations which help them attain the next higher level of language proficiency in speaking, reading, and listening comprehension. It is a 5 step process, which begins with rating and selection of reading and listening materials according to Interagency Language Roundtable (ILR) skill level descriptions. Prior to

diagnostic assessment interviews, students are asked to complete inventories about their sensory preferences, cognitive styles, and personality types. Cognitive and affective variables like extroversion-introversion, risk-taking, and differences in cognitive style (field dependence versus independence, analytic versus holistic learning) play a role in the learning process (Skehan,1989; Leaver, 1998). The three-skill interview in speaking, reading, and listening follows.

The diagnostic speaking test is similar to the Oral Proficiency Interview (OPI), which is a "personalized and individually administered elicitation and rating procedure." The examinee's foreign language communicative ability is assessed by eliciting a speech "sample rich enough to allow [for] an accurate global rating based on the ILR criteria" (Defense Language Institute Foreign Language Center Evaluation and Standardization Directorate Proficiency Standards Division, OPI 2000 Tester Certification Workshop Training Manual, 2005, pp. xi, xii). Appendix A shows descriptions of OPI tasks.

During the reading portion of the interview, examinees read several authentic texts in the target language and are asked questions about the content, specific vocabulary items, grammar structures, and idiomatic expressions demarcating socio-cultural competence. They are allowed to refer back to the texts during discussion. For the listening portion of the interview, students listen to several authentic texts in the target language, recorded from authentic sources created for the natives by the natives. There may be background noise, or speakers using regional accents. After each audio is played twice, students are asked content, vocabulary, and grammar-related questions. They are encouraged to take notes while listening.

The fundamental nature of the DA interview is to provide meaningful feedback to learners that they can understand and act upon. As Alderson claims "a crucial component of any diagnostic test must be feedback that is offered to users on their performance….. the essence of a diagnostic test must be to provide meaningful information to users which they can understand and upon which they or their teachers can act" (2006, p. 208). Feedback must be provided

as soon as possible after the test is taken since there will be little value in feedback that comes two or three weeks after the test is taken. Learners will have failed to recall how they performed and why they answered the way they did (Alderson, 2006). Due to its diagnostic nature, the DA interview aims at informing and supporting learners through the feedback it provides that is meant to be maximally educational. At the post-interview stage, students' personality and diagnostic profiles for speaking, reading, and listening are prepared. These diagnostic profiles show patterns of linguistic strengths and weaknesses. Learning plans are created outlining learning strategies, activities, resources, and recommendations for learners to follow in order to attain the next level of proficiency in each skill area. Therefore, the provision of feedback, and encouragement of learners for self-reflection are crucial. The face-to-face feedback session takes place when the DA Specialist meets with the learners to provide them with meaningful feedback based on their performance (Defense Language Institute Foreign Language Center Directorate of Continuing Education Educational Support Services Diagnostic Assessment Center Diagnostic Assessment Workshop, 2007). Follow-up sessions in subsequent weeks continue to find out whether the recommended strategies, activities, techniques or resources outlined in the learning plan are working for students or they need new sets of recommendations. Diagnostic Assessment is envisioned to integrate testing, teaching and learning by being a common denominator for all. Thus, it is "directly relevant to curricula, textbooks and other learning materials….diagnosis could be directly related to teaching and learning" (Alderson, 2006, p. 210).

This study investigates the effectiveness of diagnostic assessment as a pedagogical tool to identify the strengths and weaknesses of American language learners of Turkish in narrating past events, one of the tasks of the Oral Proficiency Interview (OPI) at Level 2. The learners are military students studying Turkish in a military institution in the United States. It focuses on improving their past narration proficiency, with greater emphasis on weaknesses that are alleviated through instruction. It discusses the importance of feedback offered on their performance, and follow-up sessions, which are crucial components of diagnostic assessment. Results of a

survey given to students illustrated students' favorable attitudes towards diagnostic assessment as a tool to improve their proficiency in past narratives.

At this military language training institution Turkish is taught in 47 weeks as a Category III language. It belongs to the 2nd category of most difficult languages to learn for English native speakers as defined by the Interagency Language Roundtable (ILR) skill level descriptions originally developed by the United States Foreign Service Institute (The ILR (FSI) Proficiency Scale, 1999). Appendix B demonstrates a complete list of categories of languages.

2. Review of Literature

Dynamic testing is at the core of the most recent form of the diagnostic assessment. It is "testing plus an instructional intervention" (Sternberg & Grigorenko, 2002, p. 23). It highlights individuals' learning potential rather than their past learning achievements. In traditional tests, occasionally called static tests, the examinees are given test items and are asked to respond to these items without feedback or intervention. Providing feedback is considered as the cause of error of measurement and thus it should be avoided. In dynamic tests, individuals are delivered the test items with explicit instruction. For instance, the examinee is given a test item to respond. If the item is answered correctly, the next item is presented. If the examinee does not respond to the item correctly, he or she is provided with clues intended to make the solution continually more obvious. The examiner then decides how many and what kinds of clues are needed for the examinee to respond to the item accurately. Coaching continues until the examinee is successful, at which time the next test item is offered (Sternberg & Grigorenko, 2002). One of the major theories underlying dynamic testing is the zone of proximal development (ZPD). "It is the distance between the actual developmental level as determined by independent problem solving and the level of potential development as determined through problem solving under adult guidance or in collaboration with more capable peers" (Vygotsky, 1978, p. 86). The zone of proximal development equips educators with an instrument with which individuals' internal course of development can be comprehended.

With this technique we can take into consideration the already established mental functions and those mental processes that are just starting to develop. Individuals' mental development can be decided by the actual developmental level and the zone of proximal development (Vygotsky, 1978).

Alderson (2006) claims that "…relatively little has been written about diagnostic testing in second and foreign language learning… and there is degree of confusion in the literature about what exactly diagnostic tests are and how they should be constructed" (p. 13). Diagnostic testing has conventionally been used in diagnosing speech and language disorders (Nation & Aram, 1984), identifying difficulties in reading and arithmetic (Clay, 1979; Bannatyne, 1971; Schonell & Schonell, 1960) and detecting learning difficulties (Bannatyne, 1971; Wilson, 1971). In other words, it has typically been used in "clinical speech and language pathology and diagnosis" ….[and] diagnostic language testing "receives only cursory treatment even in textbooks on language testing" (Alderson, 2006, p. 13). As posited by Alderson "at the heart of teaching and assessing foreign language proficiency lies the need to help learners make progress." Yet, most practitioners involved in classroom assessment procedures have neglected diagnosing foreign language learners' strengths and weaknesses in-depth (2006, p. 1).

Although diagnosis of language pathologies, reading difficulty, and learning problems have been made for many people with respect to their first language , the diagnosis of second and foreign language proficiency appears to be uncommon. It is believed that the aptitude to read in the first language is crucial to being able to partake in education, employment, and society just as disorders in an individual's first language may severely impede one's capacity to communicate as a human being. "Thus diagnosis is most developed and best understood in circumstances where the skills that are lacking or in need of remediation are important to life, education, and society, and where diagnosis can contribute to the quality of life of individuals." Speaking a second or foreign language may be crucial in multilingual societies or where an individual might be considerably underprivileged by a lack of proficiency in the language of the

mainstream culture. However, it is seldom that an individual's chances in life will be hindered "by a lack of foreign language proficiency." While the ability to read is anticipated of every child in society and inability to read is perceived as a major limitation, this is not the case for second or foreign language learning (Alderson 2006, pp. 22-23).

Consequently, diagnostic testing in the field of evaluating second or foreign language has received little consideration, and thus there seems to be "confusion and indeed ignorance" regarding what diagnostic testing is (Alderson, 2006, p. 26). As Hughes (1989) claims, there is scarcity of diagnostic tests in the twenty-first century due to its unclear and conflicting definitions in the foreign language testing literature, and lack of published research on diagnostic tests and assessment procedures.

Alderson describes on-line, computer-based diagnostic tests, DIALANG developed in 14 languages spoken in Europe, i.e., Danish, Dutch, English, Finnish, French, German, Greek, Icelandic, Irish, Italian, Norwegian, Portuguese, Spanish, and Swedish. DIALANG was developed as part of a project funded by the European Union and has tests in reading, writing, listening, grammar, and vocabulary. Test specifications were based on the Common European Framework of Reference (CEFR) (2006). Assessment of proficiency in receptive and productive skills is the focus of Common European Framework of Reference for languages (Leclercq, Edmonds & Hilton, 2014).

A list of 75 words comprises the Vocabulary Size Placement Test (VSPT) in each language. Out of these 75 words, 25 are pseudo-words. Huibregtse et al. (2002) define pseudo-words as "words that fulfill the phonological constraints of the language but do not bear meaning" (p. 227). The vocabulary items are presented in random order and test takers are not told how many non-words are included. There are VSPTs for each DIALANG language and they were created by Paul Meara and his associates of the University of Wales in Swansea (as cited in Alderson, 2006). Unlike the EFL Vocabulary Tests designed by Meara, which are graded into six frequency levels

(Tozcu & Coady, 2004; Alderson, 2006), the VSPT is not based on word frequency lists because such word frequency lists do not exist in all the DIALANG languages. The DIALANG test of Vocabulary was intended to develop tasks for the following dimensions: denotative meaning, semantic relations, combinations, and word formation (Alderson, 2006).

For the Reading Test, descriptive, narrative, expository, argumentative, and instructive texts were selected. It was ensured that objective (a fact) and subjective (opinion, attitude, mood, or wish) views are demonstrated in the selections. The test items attempted to measure the sub-skills of understanding the main idea specific details and making inferences. Topics were neither too general (to shun the testing of world knowledge) nor too specific; they were current, but not tied to a certain event to prevent the texts from ageing too rapidly (Alderson, 2006).

For the Listening Test, descriptive, narrative, expository, argumentative, and instructive text types were selected as well. Objective (a fact) and subjective (opinion, attitude, mood, or wish) views were represented in the selections. The test items on the listening test attempted to measure learners' ability to identify the main idea, specific details and making inferences (Alderson, 2006).

The DIALANG assessment system for writing aims at evaluating the communicative aspects of writing, i.e., exchanging information, expressing and finding out opinions, and persuading. The descriptive, narrative, expository, argumentative, and instructive text types were considered when constructing items to assess writing. Indirect writing items evaluated mechanics (spelling, grammar, and abbreviations), appropriacy (formal vs. informal registers), and text construction abilities (coherence and cohesion markers) (Alderson, 2006).

The DIALANG tests of grammar are sentence-based and relate to morphology and syntax. The items in the Grammar section cover core grammar structures and range from very basic to difficult.

Through different task types, learners show their ability to comprehend and produce pertinent structures (Alderson, 2006).

Testing the productive skills of direct speaking and writing is practically impossible in an on-line, computer-scored system like DIALANG. Thus, it would be highly unlikely to create computer-scored direct writing and speaking tests in particular (Alderson, 2006). He argues that there indeed is a need for diagnosing language learners' strengths and weaknesses in speaking; however, there appears to be an absence of literature and research in diagnostic assessment tests of foreign language speaking abilities and how they develop (2006).

Ehrman (1996), Leaver (1998), and Leaver, Ehrman and Shekhtman (2005) considered the importance of identifying learner differences, i.e., their learning styles, personality types, and sensory preferences, for their success in foreign language classrooms. Effective interventions can only be designed after learning difficulties are diagnosed (Ehrman, 1996). The Diagnostic Assessment Center of the Defense Language Institute Foreign Language Center (2006) designed the face-to-face Diagnostic Assessment Interview procedure for the purpose of identifying learners' strengths and weaknesses in speaking, reading, and listening after discovering their learning styles through questionnaires about cognitive styles, personality types, and sensory preferences. Focusing on the strengths facilitates the identification of the learner's current level and focusing on the weaknesses leads to taking remedial measures through further instruction. The diagnostic assessment interview allows for a comprehensive analysis of learner responses to tasks, and provides detailed feedback immediately after the interview. The interview results have suggestions for further learning or instruction. The diagnostic assessment interview for speaking used in this study identified learners' strengths and weaknesses in both their knowledge and use of language in narrating past events.

3. Methods

3.1 Purpose of Study

The research attempted to answer the following questions: (1) Do the students in an intensive Turkish program who are given Diagnostic Assessment Interviews use more cohesive ties in narrating past events than those in a control group? (2) Do the students who are given Diagnostic Assessment Interviews use cohesive ties more accurately in narrating past events as compared to the control group? (3) Do the students who are given Diagnostic Assessment Interviews exhibit significantly better control over basic grammar structures in narrating past events than a control group?

3.2 Theoretical Framework

The present study adopts and explores the rationale of the theory of formative evaluation of a learner's skills, strengths, and weaknesses proposed by Leaver (1998). The assumption behind this theory is that specific areas of work a student needs to exceed should be indicated in advance. Learners should be provided feedback and recommendations which help them attain the next higher level of language proficiency in all skill areas. Prior to diagnostic assessment interviews, students are asked to complete inventories about their sensory preferences, cognitive styles and personality types. There is emphasis on the role of cognitive and affective variables i.e., extroversion-introversion, differences in cognitive style (field dependence versus independence, analytic versus holistic learning) and sensory preferences (Skehan, 1989; Leaver, 1998).

3.3 Participants

Participants of this study were 24 military students enrolled in an intensive Turkish language program in the United States. As they were entering the program, all participants took the Defense Language Aptitude Battery (DLAB), which measured their aptitude and determined their placement in the Turkish program. Students were randomly assigned to treatment and control groups . The mean of participants' age was 33. Nineteen students were male and five were female.

3.4 Description of Narration

According to Labov, the study of narrative covers a wide range of human experiences, i.e., novels, short stories, film, folk tale, interviews, oral memoirs, chronicles, histories, comic strips, and other visual media. These forms of information transmission may be driven from the deep-seated human ability to transfer experience from one person to another by means of oral narratives of personal experiences. The progress in sociolinguistic research in 1960s instigated spontaneous recounting of personal experiences (To appear in the Cambridge Encyclopedia of the Language Sciences).

Narratives are recounted with a comparable organization in a wide variety of societies and cultures. A narrative as defined by Labov is one way of telling past events whereby the sequence of the narrative clauses corresponds to the order of events as they happened, which is often denoted as temporal organization. Each narrative clause answers an impending question: "what happened then?" (To appear in the Cambridge Encyclopedia of the Language Sciences). Chronology or temporal arrangement of events in a story, entails "that each unit of meaning in the discourse be located in logical and proper relation to the other units of meaning that precede or follow it. Narration, for example, demands a cohesiveness that most often rests on chronology, an ordered sequencing of all components connected……and organization within a logical textual framework, or in other words, in a paragraph: the characteristic text type of Level 2" (Defense Language Institute Foreign Language Center Evaluation and Standardization Directorate Proficiency Standards Division, OPI 2000 Tester Certification Workshop Training Manual, 2010, p. 146). Cohesive devices "are words and phrases that link ideas and move the action forward in some form of logical narrative order…Adverbs and conjunctions serve most frequently as cohesive devices…They permit logical sequencing, create structures of meaning by establishing time frames for actions and events, and help create and sustain comparisons and contrasts" (Defense Language Institute Foreign Language Center Evaluation and Standardization Directorate Proficiency Standards Division, OPI 2000 Tester Certification Workshop Training Manual, 2010, p. 145). Adept speakers are able to explicate intricate matters in detail and

deliver extensive and coherent narration. They produce clear, well-organized, detailed narrations on complex themes (as cited in Hilton, 2014). Proficient narrative discourse encompasses accurate use of cohesive devices and grammar constructions (Forsberg & Bartning, 2010; Wei, 2011).

3.5 Description of Pre-interview Materials

3.5.1 Barsch Learning Style Inventory

This inventory is a quick evaluation of a student's learning style preferences, i.e., sensory modalities as defined by Ehrman (1996). It comprises 32 statements with quantified values used in the scoring procedure. Answers are based on students' actual sensory preference, physiological modalities through which they identify and grasp incoming new information and not areas which they would like to have as strengths (Western Oregon University Learning Center). This inventory was used as a tool to determine the sensory preferences of participants by categorizing them into three groups: Primary Visual Learners who are impacted by information which they can see and learn best by seeing; Primary Auditory Learners who are influenced by information to which they can listen and learn best through discussions; and Primary Tactile/Kinesthetic Learners who learn best through small and large motor movements, i.e., moving, touching, and discovering the physical world around them through a hands-on approach. Appendix C shows the descriptions of sensory preferences.

3.5.2 The E&L Construct for Cognitive Styles

Cognitive styles are routine patterns of information processing (Leaver, Ehrman & Shekhtman, 2005). They shape an individual's way of perceiving, processing, and organizing the world (Leaver, 1998). Ehrman and Leaver (1997, 2003) chose different "cognitive style scales" and organized them into a model based on two primary categories. They were named synopsis (adjective-synoptic) referring to holistic, and ectasis (adjective – ectenic) referring to atomistic. In foreign language education, synoptic learning involves intuition and subconscious control of the learner whereas ectenic learning

transpires under conscious control. Each "pole" consists of "ten cognitive scales which are subscales in the E&L Construct" (Leaver, Ehrman & Shekhtman, 2005, p. 70) as illustrated in Appendix D. Synoptic and ectenic learning constitute the umbrella scale encompassing subscales that emulate various facets of synoptic and ectenic learning. Each scale is a continuum; an individual would be more or less global or particular, for instance, not entirely global or entirely particular. A student's scores for each subscale on the E & L construct show a unique assortment of ten different styles, which is commonly referenced as a learning profile. In a classroom, each student will have a very different or slightly different learning profile, showing details about their approach to foreign language learning. These profiles specify in advance where the student may have a deviance from the textbook, teacher, or the teaching methodology in the language program. Identifying where these deviations occur means that the student can build strategies for the opposite, required style (Leaver, Ehrman, and Shekhtman, 2005).

Ehrman (1996) postulates that learning strategies are specific techniques or activities that students employ to learn. According to Leaver, Ehrman, and Shekhtman, some strategies are utilized consciously whereas others are automatic. Most learning styles are manifested by observable learning strategy behaviors. The ability to choose the appropriate learning strategies for learning is called strategic competence. Using learning strategies effectively is crucial for the learning process. Monitoring learners' performance in particular tasks may ease the provision of feedback by the teacher on strategy use. Learners should have some flexibility in the strategies they use; if they continue using the same strategies, their learning may be fairly inadequate (2005). Learning styles may be "flexed" by familiarizing learners with a wide range of strategies and coaching them in using these strategies during specific tasks (Defense Language Institute Foreign Language Center Directorate of Continuing Education Educational Support Services Diagnostic Assessment Center, Diagnostic Assessment Workshop, 2007). "…Although most learning style models are bipolar (i.e., they have two clearly established end points), they really represent a continuum of behavior" (Ehrman, 1996, p. 51). These preferences, which may

run an array from mild to strong, can be described as "comfort zones," with which individuals find themselves at ease, but can do in a different way if required (p. 54).

3.5.2 Myers Briggs Personality Type Indicator

Personality typing through Myers-Briggs type indicator (MBTI) (Myers with Myers, 1980; Myers, McCaulley, et al., 1998) and the Keirsey Temperament Sorter (Keirsey & Bates, 1978; Keirsey, 1998) have been used extensively for educational purposes. They were driven from the work of Swiss psychiatrist Carl Jung (1971). Dimensions of mental activity are defined as 1) extraversion versus introversion (direction of energy flow); 2) sensing and intuition (mental function for absorbing data); 3) thinking versus feeling (mental function for reaching conclusions and making decisions); and 4) judging versus perceiving (preference to deal with the outside world, particularly how much structure an individual wants from it) (as cited in Leaver, Ehrman & Shekhtman, 2005). According to Jung, one member from each dimension is generally preferred and consciously used. The other member stays with the individual as well, however, it impacts unconscious functioning more (Jung, 1971). There are sixteen possible amalgamations of these four domains, which are called personality types and none of these sixteen types is considered better than the other; however, some settings deliver a better fit for some personality types than for others (Leaver, 1998; Leaver, Ehrman, & Shekhtman, 2005). The combination of these four domains into 16 personality types and their implications for the classroom are summarized in Appendix E.

4. Procedures

In an effort to determine the effect of diagnostic assessment on the level of proficiency students have attained in narrating past events and find out whether there has been further improvement in their narrative abilities, participants in both the treatment and control groups were given diagnostic assessment interviews as pre-and posttests; one interview in the middle of the second semester (week 28) and another in the middle of the third semester (week 38). The elicitation for the Past narration task at the pre-test level was how

they spent Christmas. At the post-test level, they were asked to narrate their most recent trip. During the Pre-interview stage, participants were asked to complete inventories related to sensory preferences (Barsch Learning Style Inventory), cognitive styles (the E & L Construct for Cognitive Styles), and personality types (Myers Briggs Personality Type Indicator). The students in the treatment group were given comprehensive feedback as to their strength and weaknesses through speaking profiles. They were provided with a personalized learning plan with recommendations to improve upon their weaknesses demonstrated during the DA interview. The students in the control group, on the other hand, were required to read two 2-page narratives per week, and answer four comprehension questions based on the narratives. The questions required short answers. They were also asked to prepare a single page summary for each narrative. They were instructed to pay attention to the organization of the narratives and grammar structures embedded in them. They were provided with standard teacher feedback for improvement. In sum, the control students performed reading and reading comprehension activities similar to those in class. Other than reading these narratives the participants in the control group received the same type of regular instruction in Turkish that they were already receiving together with the treatment group. At the conclusion of the interview students were given a questionnaire about strategies they had been using to learn the target language grammar, vocabulary, and to enhance their overall speaking ability including their past narration skills, as shown in Appendix F. One month after the study, students were given a survey which asked for their reaction to the diagnostic assessment interview and personalized learning plan. Table 1 demonstrates students' favorable attitudes to diagnostic assessment interview.

The DA interview helped me use person markers accurately	100%
The DA interview helped me use case markers accurately	100%
The DA interview helped me use pluralization markers accurately	100%
The DA interview helped me use passive voice accurately	95%
The DA interview helped me use nominalizations accurately	95%

The DA interview helped me use more cohesive devices	100%
The DA interview helped me use cohesive devices more accurately	100%

Table 1. Student Survey: The percentage of those who answered agree or strongly agree

5. Results and Discussion

The results indicate that the students who were given diagnostic assessment interviews did experience a significant improvement in using person, case, pluralization, passive, nominalization markers and cohesive devices. Actually, both the control and treatment groups showed increases. However, the students in the treatment group showed significantly greater gains than the ones in the control group.

		Control		Treatment	
		M	SD	M	SD
Person Markers	Pretest	2.25	.45	2.00	.60
	Posttest	3.75	.62	7.17	.94
	Gain (Post-Pre)	1.5		5.17	
Case Markers	Pretest	6.00	1.04	6.33	1.07
	Posttest	9.92	1.16	16.08	1.08
	Gain (Post-Pre)	3.92		9.75	
Pluralization Markers	Pretest	4.08	.51	4.25	.45
	Posttest	6.25	.87	9.42	.90
	Gain (Post-Pre)	2.17		5.17	
Passive Markers	Pretest	1.00	.00	1.00	.00
	Posttest	1.50	.52	2.92	.67
	Gain (Post-Pre)	0.50		1.92	
Nominalization Markers	Pretest	1.25	.45	1.33	.49
	Posttest	1.50	.52	3.33	.65
	Gain (Post-Pre)	0.25		2.00	
Number of Cohesive Devices	Pretest	3.08	.79	3.25	.62
	Posttest	4.42	.79	8.42	1.16
	Gain (Post-Pre)	1.34		5.17	
Number of Accurately	Pretest	2.08	.51	2.00	.43

Used Cohesive Devices	Posttest	3.50	.52	7.33	1.07
	Gain (Post-Pre)	1.42		5.33	

Table 2. Means and standard deviations for person markers, case markers, pluralization markers, passive markers, nominalization markers, cohesive devices and accurately used cohesive devices for Treatment and Control Groups

The means and standard deviations for person markers for Treatment and Control groups are presented in Table 2. The data were analyzed using repeated measures ANOVAs using Group (Treatment vs. Control) and Time (Pre-test vs. Post-test) using Group as the independent variable with repeated measures on Time. The main effect of Time was found significant. $F (1, 22) = 550.00$, $p<.001$. The Group by Time interaction was also found significant. $F(1, 22) = 266.20$, $p<.001$. The main effect of Group was found to be significant as well. $F (1, 22) = 84.49$, $p<.001$. The significant interaction indicates a differential improvement in person marker use for the Treatment and Control groups. This significant interaction effect was further investigated by using two t-tests with a Bonferroni correction with alpha levels adjusted to .005 (.01/2=.005). The results illustrate that there was no statistically significant difference in person marker usage between the Treatment and Control students at the pre-test level [$t(22)=1.15$, $p>.05$] whereas there was a significant difference between the two groups at the post-test level [$t(22)=-10.52$, $p<.001$]. In other words, both the Treatment group and Control group were similar in person marker usage at the beginning of the study. However, the gap between the two groups became larger at the end of the study.

The means and standard deviations for case markers for Treatment and Control groups are presented in Table 2. The data were analyzed using repeated measures ANOVAs using Group (Treatment vs. Control) and Time (Pretest vs. Posttest) using Group as the independent variable with repeated measures on Time. The main effect of Time was found significant. $F (1, 22) = 577.84$, $p<.001$. The Group by Time interaction was also found significant. $F(1, 22) = 286.70$, $p<.001$. The main effect of Group was found to

be significant as well. F (1, 22)= 150.73, p<.001]. The significant interaction indicates a differential improvement in case marker use for the Treatment and Control groups. This significant interaction effect was further investigated by using two t-tests with a Bonferroni correction with alpha levels adjusted to .005 (.01/2=.005). The results illustrate that there was no statistically significant difference in case marker use between the Treatment and Control students at the pre-test level [t(22)=-.77, p>.05], whereas there was a significant difference between the two groups at the post-test level [t(22)=-13.43, p<.001]. In other words, both the Treatment group and Control group were similar in their case marker use at the beginning of the study. However, the gap between the two groups became larger at the end of the study.

The means and standard deviations for pluralization markers for Treatment and Control groups are presented in Table 2. The data were analyzed using repeated measures ANOVAs using Group (Treatment vs. Control) and Time (Pretest vs. Posttest) using Group as the independent variable with repeated measures on Time. The main effect of Time was found significant. F(1, 22) = 287.78, p<.001. The Group by Time interaction was also found significant. F(1, 22)= 118.80, p<.001. The main effect of Group was found to be significant as well. F(1, 22)= 59.46, p<.001. The significant interaction indicates a differential improvement in pluralization marker use for the treatment and control groups. This significant interaction effect was further investigated by using two t-tests with a Bonferroni correction with alpha levels adjusted to .005 (.01/2=.005). The results illustrate that there was no statistically significant difference in pluralization marker use between the Treatment and Control students at the pre-test level [t(22)=-.84, p> .05] whereas there was a significant difference between the two groups at the post-test level [t(22)= -8.78, p<.001]. In other words, both the Treatment group and Control group were similar in their pluralization marker use at the beginning of the study. However, the gap between the two groups became larger at the end of the study.

The means and standard deviations for passive markers for Treatment and Control groups are presented in Table 2. The data were analyzed using repeated measures ANOVAs using Group

(Treatment vs. Control) and Time (Pre-test vs. Post-test) using Group as the independent variable with repeated measures on Time. The main effect of Time was found significant. $F(1,22) = 111.46$, $p<.001$. The Group by Time interaction was also found significant. $F(1, 22) = 29.71$, $p<.001$. The main effect of Group was found to be significant as well. $F(1, 22) = 29.71$ $p<.001$. The significant interaction indicates a differential improvement in Passive marker use for the Treatment and Control groups. This significant interaction effect was further investigated by using two t-tests with a Bonferroni correction with alpha levels adjusted to .005 (.01/2=.005). The results illustrate that there was no difference in passive marker use between the Treatment and Control students at the pre-test level whereas there was a significant difference between the two groups at the post-test level [$t(22) = -5.79$, $p<.001$]. In other words, both the Treatment group and Control group were similar in their passive marker use at the beginning of the study. However, the gap between the two groups became larger at the end of the study.

The means and standard deviations for nominalizations for Treatment and Control groups are presented in Table 2. The data were analyzed using repeated measures ANOVAs using Group (Treatment vs. Control) and Time (Pre-test vs. Post-test) using Group as the independent variable with repeated measures on Time. The main effect of Time was found significant. $F(1, 22) = 157.24$, $p<.001$. The Group by Time interaction was also found significant. $F(1, 22) = 95.12$, $p<.001$. The main effect of Group was found to be significant as well. $F(1, 22) = 16.77$, $p<.05$. The significant interaction indicates a differential improvement in nominalizations for the Treatment and Control groups. This significant interaction effect was further investigated by using two t-tests with a Bonferroni correction with alpha levels adjusted to .005 (.01/2=.005). The results illustrate that there was no statistically significant difference in nominalization use between the Treatment and Control students at the pre-test level [$t(22) = -.43$, $p>.05$] whereas there was a significant difference between the two groups at the post-test level [$t(22) = -7.61$, $p<.001$]. In other words, both the Treatment group and Control group were similar in their use of nominalizations at the

beginning of the study. However, the gap between the two groups became larger at the end of the study.

The means and standard deviations for cohesive devices for Treatment and Control groups are presented in Table 2. The data were analyzed using repeated measures ANOVAs using Group (Treatment vs. Control) and Time (Pre-test vs. Post-test) using Group as the independent variable with repeated measures on Time. The main effect of Time was found significant. $F(1, 22) = 265.57$, $p < .001$. The Group by Time interaction was also found significant. $F(1, 22) = 98.63$, $p < .001$. The main effect of Group was found to be significant as well. $F(1, 22) = 72.37$, $p < .001$.The significant interaction indicates a differential improvement in cohesive device use by the Treatment and Control students. This significant interaction effect was further investigated by using two t-tests with a Bonferroni correction with alpha levels adjusted to .005 (.01/2=.005). The results illustrate that there was no statistically significant difference in the number of cohesive devices used by the Treatment and Control students at the pre-test level [$t(22) = -.57$, $p > .05$] whereas there was a significant difference between the two groups at the post-test level [$t(22) = -9.84$, $p < .001$]. In other words, both the Treatment group and Control group were similar in the number of cohesive devices they used at the beginning of the study. However, the gap between the two groups became larger at the end of the study.

The means and standard deviations for accurately used cohesive devices for Treatment and Control groups are presented in Table 2. The data were analyzed using repeated measures ANOVAs using Group (Treatment vs. Control) and Time (Pre-test vs. Post-test) using Group as the independent variable with repeated measures on Time. The main effect of Time was found significant. $F(1, 22) = 370.11$, $p < .001$. The Group by Time interaction was also found significant. $F(1, 22) = 135.75$, $p < .001$. The main effect of Group was found to be significant as well. $F(1, 22) = 101.71$ $p < .001$. The significant interaction indicates a differential improvement in accurately used cohesive devices by the Treatment and Control groups. This significant interaction effect was further investigated by using two t-tests with a Bonferroni correction with alpha levels

adjusted to .005(.01/2=.005). The results illustrate that there was no statistically significant difference in the number of accurately used cohesive devices by the Treatment and Control students at the pre-test level [$t(22) = .43$, $p > .05$ whereas there was a significant difference between the two groups at the post-test level [$t(22) = -11.13$, $p < .001$]. In other words, both the Treatment group and Control group were similar in their accurate use of cohesive devices at the beginning of the study. However, the gap between the two groups became larger at the end of the study.

6. Conclusion

The theoretical framework underlying this study was formulated based on the theory of formative evaluation of a learner's skills, strengths, and weaknesses proposed by Leaver (1998). The assumption behind this theory is that specific areas of work a student needs to improve on should be indicated in advance. Learners should be provided feedback and recommendations which help them attain the next higher level of language proficiency in all skill areas. The data demonstrated that diagnostic assessment interviews do result in a significant improvement in using person, case, pluralization, passive, nominalization markers, and cohesive devices.

Although both the treatment and control groups showed increases in using person, case, pluralization, passive, nominalization markers, and cohesive devices (presumably because they were both studying Turkish full time), the students in the treatment group showed significantly greater gains than the ones in the control group. They narrated a series of events with detail in a logical and chronological order by using cohesive devices in paragraph long discourse. They placed events in context and made the distinction between the more important and less important details in the narration. They also demonstrated to have better control over basic grammatical structures.

6.1 Pedagogical Implications

The findings of this study indicate that diagnostic assessment interviews will almost certainly facilitate proficiency in basic grammar structures, i.e., person, case, pluralization, passive, nominalization markers, and cohesive devices. This study indicates clear and positive findings in support of such a pedagogical method because a large benefit was gained for a rather small amount of time. Furthermore, diagnostic assessment is beneficial since it is done outside of class without utilizing precious classroom time.

This study enables teachers and practitioners in the field to create meaningful opportunities for "diagnostic" teaching in the classroom. This can complement the practice of conducting formal diagnostic assessment interviews outside of class.

6.2 Limitations of the Study and Implications for Future Research

The significance of this study lies in its experimental nature as there is not much research done in this area. The data demonstrated that diagnostic assessment interviews do result in a significant increase in learner proficiency in narrating past events. The following issues warrant further investigation. The present study is limited in scope in that it did not include a greater sample of participants with a wider scale of proficiency levels. It would be worthwhile to do so. The same study with a greater number of participants with a wider range of proficiency levels. This experiment focused on the basic sentence structures in Past Narration. An alternative approach would be to examine the participants' performance on more complex grammar features (i.e., past, present and future conditionals, etc.), both before and after such study and attempt to determine what type of contribution is being made by this approach to overall proficiency in grammar. Furthermore, it would be worth exploring the effectiveness of diagnostic assessment on student performance in a variety of OPI tasks, (i.e., Present Narration, Future Narration, Instructions, Descriptions, Supported Opinion, etc.) in Turkish and other languages taught at Defense Language Institute Foreign

Language Center. Finally, a longitudinal study could be conducted in order to determine the long-term effects of diagnostic assessment on basic and complex grammar structures. The aforementioned expansions to this study can pave the way to more comprehensive research in diagnostic assessment.

Note:
The views, opinions, and or findings contained in this report are those of the author and should not be construed as an official Department of the US Army position, policy, or decision unless designated by other official documentation.

Acknowledgement

I would like to thank Dr. Yu Chang Chen, Educational Researcher in the Training and Analysis Division at DLIFLC for assisting me with the statistical analysis of the data.

References

Alderson, C. (2006). *Diagnosing foreign language proficiency: The interface between learning and assessment*. New York: Continuum.

Bandler, R. (1985). *Using your brain—for a change: neurolinguistic programming*. Moab, UT: Real People Press.

Bannatyne, A. (1971). *Language, reading and learning disabilities: psychology, neuropsychology, diagnosis and remediation*. Springfield, IL: Charles C. Thomas Publisher.

Clay, M. M. (1979). *The early detection of reading difficulties: a diagnostic survey with recovery procedures.* Auckland: Heinemann.

Defense Language Institute Foreign Language Center Continuing Education. (2006). *Diagnostic assessment workshop training manual.* Presidio of Monterey, CA: Author.

Defense Language Institute Foreign Language Center Directorate of Continuing Education Educational Support Services Diagnostic Assessment Center, Diagnostic Assessment Workshop. (2007). *Diagnostic assessment: A tool for learners, teachers, and program managers*. Presidio of Monterey, CA: Author

Defense Language Institute Foreign Language Center Evaluation and Standardization Directorate Proficiency Standards Division. (2005). *OPI 2000 tester certification workshop training manual.* Presidio of Monterey, CA: Author.

Defense Language Institute Foreign Language Center Evaluation and Standardization Directorate Proficiency Standards Division. (2010). *OPI 2000 tester certification workshop training manual.* Presidio of Monterey, CA: Author.

Ehrman, M. E. (1996). *Understanding second language learning difficulties*. Thousand Oaks, CA: Sage Publications.

Ehrman, M. E., & Leaver, B. L. (1997). Sorting out global and analytic functions. Orlando, FL: AAAL Annual Meeting.

Ehrman, M. E., & Leaver, B. L. (2003). Cognitive styles in the service of language learning. *System,* 31, 393-415.

Forsberg, F., & Bartning, I. (2010). Can linguistic features discriminate between the communicative CEFR- levels? a pilot study of written L2 French. In I. Bartning, M. Martin & I. Vedder (eds.) *Communicative proficiency and linguistic*

development (pp. 133-158). EUROSLA Monograph Series. Retrieved 23 September 2012 from http://eurosla.org/monographs/EM01/EM01home.php

Gregorc, A. (1982). Learning style/brain research: harbinger of an emerging psychology. In J. Keefe (ed.), *Student learning styles and brain behavior*. Reston, VA: National Association of High School Principals.

Hilton, H. (2014). Oral fluency and spoken proficiency: considerations for research and testing. In P. Leclercq, A. Edmonds & H. Hilton (eds.) *Measuring L2 proficiency: perspectives from SLA* (pp. 27-53). Toronto: Multilingual Matters.

Holzman, P. S., & Gardner, R. W. (1959). Leveling-sharpening and memory organization. *Journal of Abnormal and Social Psychology*, 61, 176-180.

Huibregtse, I., Admiraal, W., & Meara, P. (2002). Scores on a yes-no vocabulary test: correction for guessing and response style. *Language Testing*, 19, 3, 227-45.

Hughes, A. (1989). *Testing for language teachers*. Cambridge: Cambridge University Press.

Jung, C. G. (1971). Psychological types (H. G. Baynes trans., rev. R. F. Hull). Reprinted 1971. In Sir H. Read, M. Fordham, G. Adler & W. McGuire (eds.), *Collected works of C. G. Jung Vol. 6, Bollingen Series 20*. Princeton, NJ: Princeton University (Original work published 1921).

Kant, I. (1781/1998). *The critique of pure reason*, trans. P. Guyer and A. W. Wood. Cambridge: Cambridge University Press.

Keirsey, D. (1998). *Please understand me II: temperament, character, intelligence*. Del Mar, CA: Prometheus Nemesis Book Company.

Keirsey, D., & Bates, M. (1978). *Please understand me: Character and Temperament Types*. Del Mar, CA: Prometheus Nemesis Book Company.

Labov, W. *Oral narratives of personal experience*. (Forthcoming in the Cambridge Encyclopedia of Language Sciences). Retrieved June 4 2010 from www.ling.upenn.edu/~w**labov**/Papers/Feb**Oral**NarPE.pdf

Leaver, B. L. (1998). *Teaching the whole class* (5th Ed.). Dubuque, Iowa: Kendall/Hunt Publishing Company.

Leaver, B. L., Ehrman, M., & Shekhtman, B. (2005). *Achieving success in second language acquisition*. Cambridge: Cambridge University Press.

Leaver, B. L., & Campbell, C. (2015). Experience with higher levels of proficiency. In T. Brown & J. Bown (eds.) *To advanced proficiency and beyond: theory and methods for developing superior second language ability* (pp. 3-21). Washington, D.C.: Georgetown University Press.

Leclercq, P., Edmonds, A., & Hilton, H. (eds.) (2014). *Measuring L2 proficiency: perspectives from SLA*. Toronto: Multilingual Matters.

Messick, S. (1984). The nature of cognitive styles: problems and promise in educational practice. *Educational Psychologist*, 59, 59-74.

Myers, I. B. with P. Myers (1980). *Gifts differing*. Palo Alto, CA: Consulting Psychologists.

Myers, I. B., McCaulley, M. H., Quenk, N. L., & Hammer, A. L. (1998). *MBTI manual: a guide to the development and use of the Myers-Briggs Type Indicator* (third edition). Palo Alto, CA: Consulting Psychologists.

Nation, J. E., & Aram, D. M. (1984). *Diagnosis of speech and language disorders* (second edition). San Diego, CA: College-Hill Press.

Peirce, C. S. (1878). Deduction, induction, and hypothesis. *Popular Science Monthly*, 13, 470-482.

Pimsleur, P. R., & Struth, J. F. (1968). Knowing your students in advance. *Modern Language Journal*, 53, 2, 85-87.

Ramírez III, Manuel, & Castañeda, A. (1974). *Cultural democracy, bicognitive development and education*. NY: Academic.

Reed, D. (2006). Diagnosing assessment in language teaching and learning. *Clear News*, 10, 2, 1-7.

Schonell, F. J., & Schonell F. E. (1960). *Diagnostic and attainment testing, including a manual of tests, their nature, use, recording and interpretation.* Edinburgh: Oliver and Boyd.

Skehan, P. (1989). *Individual differences in second-language learning.* London: Edward Arnold.

Sternberg, R. J., & Grigorenko, E. L. (2002). *Dynamic testing: the nature and measurement of learning potential.* Cambridge: Cambridge University Press.

The ILR (FSI) proficiency scale. (1999). Retrieved 12 December, 2007 from http://www.sil.org/ lingualinks/languagelearning/mangngyryrlngglrnngprgrm/ TheILRFSIPpoficiencyScale.htm.

Tozcu, A., & Coady, J. (2004). Successful learning of frequent vocabulary through CALL also benefits reading comprehension and speed. *Computer Assisted Language Learning*, 17, 5, 473-495.

Vygotsky, L. S. (1978). *Mind in society: the development of higher psychological processes.* (M. Cole, V. John-Steiner, S. Scribner & E. Souberman, eds.) Cambridge, MA: Harvard University Press.

Wei, M. (2011). Investigating the oral proficiency of English learners in China: a comparative study of the use of pragmatic markers. *Journal of Pragmatics,* 43, 3455-3472.

Western Oregon University Learning Center. *Barsch Learning Style Inventory*. Retrieved July 8 2010 from www.wou.edu/provost/aalc/learning/barsch_ls__ inventory.php

Witkin, H. A., & Goodenough, D. R. (1981). *Cognitive styles: essence and origins: Field dependence and field independence.* NY: International Universities.

Wilson, A. (ed.) (1971). *Diagnosis of learning difficulties*. New York: McGraw-Hill Inc.

Appendices

Appendix A: Description of OPI Tasks

Level	Task	Description
1	**Simple Short Conversations**	"should focus on topics such a favorite foods and drinks, hobbies, background information (including age, nationality, and profession), family and friends, and other topics in the Level 1 domain."
1	**Examinee Asks Questions**	"…testers must provide a suitable context appropriate for Level 1 and ask the examinee in English (because of examinee's low listening skill in the target language) to pose questions."
1	**Role Play (Survival Situation)**	"The tester starts the Role-Play, but only after making certain the examinee understands the situation…. Some typical Level 1 Role-Plays might task the examinee to get a room in hotel, make a dinner reservation, buy a train ticket…."
2	**Narration** **Past, Present, Future**	"…At Level 2, narration consists of relating a series of events in a logical and chronological order. Level 2 speakers are able to tell stories in connected discourse of paragraph length. They are able to place events in context and provide some detail. An important feature of the narration is the ability to make the distinction between the more important and less important matters in the story. Speakers at Level 2 are able to narrate in all three major time frames: past, present, and future. They must be able to use adverbs and conjunctions appropriately to indicate sequence and link the utterances in paragraph-length discourse… The content areas are the examinees' own background and everyday life…."

2	**Description**	"….as an OPI task means concrete, physical description, where the examinee talks about what a person, place or thing looks like. Level 2 speakers demonstrate the ability to produce a large number of appropriate qualifiers (adjectives, prepositional/ postpositional phrases, descriptive dependent clauses), as well as vocabulary that is detailed enough to form a mental picture in the mind of the tester by means of organized discourse."
2	**Instructions or Directions**	"….the examinee must demonstrate the ability to give detailed and extensive instructions or directions….In eliciting instructions, testers should choose appropriate topics that require the use of common, high-frequency vocabulary, and not technical terms generally unfamiliar to Level 2 speakers. Testers should ask questions about how to prepare to do something, not how to actually do something such as operate a machine or an appliance…In eliciting directions from one place to another, the tester must specify not only the destination, but also the starting point….Directions must be organized so that the tester can follow a logical sequence."
2	**Reporting Facts**	"…Examinees must be able to demonstrate that they can report facts about well-known current events (international, national, or local) that have been widely reported in the media."
2	**Situation with a Complication**	"Routine situations are the predictable components of daily life, and are standardized to the extent that all people are engaged in them on a regular basis. They include "survival" situations (e.g. obtaining food and shelter, arranging transportation), social needs and exchange of goods and services, and habitual activities that form the context of an

		individual's life (e.g. school, work, home, friends, and pastimes). Routine situations become non-routine with the introduction of complications, which add an unexpected twist. These complications require….. providing explanations and giving reasons….. Thus, ordering a meal, a routine situation, would require answering and asking questions using discourse at the sentence level (Basic Survival Situation). On the other hand, trying to pay for a meal with an expired credit card (Situation with a Complication) would require an explanation such as how the customer proposes to resolve the situation."
3	**Supported Opinion**	"….requires abstract vocabulary and linguistic formulations. This means that they not only can talk about people, places and experiences, but can refer to, and discuss, significant social or political issues and developments…speakers may well include personal references, but always with the purpose of providing evidence to support a point of view."
3	**Hypothesis**	"…is a proposal of the "what if" and "under what circumstances" rather than a statement of the "what is" of everyday reality. …. Hypothesis calls for abstract speculation at the societal level that moves beyond the concrete and personal discourse of Level 2…"
3	**Discuss an Abstract Topic**	"…calls for elaboration of a social, political or a similar issue at a higher level, not necessarily the examinee's opinion. The abstract approach may involve deliberation in which competing ideas are discussed. In so doing, it involves the use of language with suitably abstract formulations."

3	**Unfamiliar Situation**	"…Examinees must be able to deal with a non-routine problem that they may encounter in the target culture, such as talking one's way out of a traffic ticket.… [Unfamiliar] situations involve people with whom the examinee does not have a close relationship."
4	**Role-Play Situations**	"..two Role-Plays must be given to check the examinees' use of both *formal* and *informal* register…."
4/5	**Supported Opinion** **Hypothesis** **Discuss an Abstract Topic**	"…Level 4 is the highest test possible in the OPI testing system. Performance beyond Level 4 is determined by the quality of the language sample…A Level 4 speaker differs from a Level 5 speaker in that they cannot sustain the sophisticated, flexible, native-like performance at Level 5. The Level 4 speaker can be "slightly off" in intonation, stress, accent, use of idioms, or cultural references….Three tasks used at Level 3, Supported Opinion, Hypothesis, and Abstract Topic, are also required tasks at Level 4. Descriptive Preludes at Level 4 differ, however, from those at Level 3, in that they go beyond Level 3 societal topics to address highly abstract moral, conceptual and philosophical questions. Testers need to raise their language to elicit sophisticated responses in extended discourse…"
4/5	**High Level Colloquialisms and Proverbs**	"…socio-cultural elements of the language, such as high level colloquialisms and proverbs, must be included in the test. This type of task is used to determine if the examinee can be rated higher than Level 4… Examinees can be asked either to explain sayings or to complete the second half of sayings and explain…"

4/5	**Information Pass**	"..is another task for testing at Level 4 and above to determine if the examinee can perform higher than 4. …..Information Pass requires the examinee to spontaneously convert an informal to a formal message and vice-versa…"

Adapted from Defense Language Institute Foreign Language Center Evaluation and Standardization Directorate Proficiency Standards Division, OPI 2000 Tester Certification Workshop Training Manual, 2010, pp. 131- 178.

Appendix B: Categories of Languages

Category IV: The category of most difficult languages to learn for English native speakers	Chinese Korean Japanese Arabic Pashto
Category III: The second category of most difficult languages to learn for English native speakers	Turkish Armenian Russian Dari Farsi Hindi Urdu
Category II: The second category of easiest languages to learn for English native speakers	German Indonesian
Category I: The category of easiest languages to learn for English native speakers	Spanish Portuguese French Italian Norwegian

(The ILR (FSI) Proficiency Scale, 1999)

Appendix C: Sensory Preferences

Visual Learners	"**Visual** learners acquire new information through sight. Distinctions that are important to visual learners include brightness, size, color, saturation, distance, clarity, contrast, texture, frame, and symmetry (Bandler, 1985). Visual learners can be subdivided into two groups: **verbalist** (they see words) and **imagist** (they see pictures)."
Auditory Learners	"**Auditory** learners acquire new information through sound. Distinctions that are important to them include pitch, tempo, rhythm, timbre, and resonance (Bandler, 1985). Auditory learners can be further divided into two groups: **aural** (they learn by listening to others) or **oral** (they learn by talking and hearing themselves)."
Motor Learners	"Motor learners acquire new information through movement. Distinctions that are important to them include frequency, pressure, duration, and intensity (Bandler, 1985). Motor learners can be subdivided into two groups: **kinesthetic** (they learn through the use of gross motor muscles) or **mechanical** (they learn through the use of fine motor muscles)."

(Adapted from Leaver, 1998, pp. 25-26)

Appendix D: The E&L Construct for Cognitive Styles

synoptic learning	*(definitions)*	**ectenic learning**	*(definitions)*	category source
analogue	learning through metaphor	**digital**	literal and factual learning	Ehrman and Leaver
concrete	hands-on learning	**abstract**	learning through ideas and books	Gregorc
field independent	decontextualized learning	**field dependent**	Contextuali -zed	Witkin and Goodenough

			learning	
field sensitive	learning through osmosis	**field insensitive**	lack of osmosis in learning	Ehrman; Ramírez and Castañeda
global	oriented toward the big picture	**particular**	oriented toward details	Ehrman and Leaver
impulsive	simultaneous thought and reaction	**reflective**	reaction following thought	Messic
inductive	understanding rules from examining examples	**deductive**	learning rules, then understand-ing examp-les	Pierce
leveling	noticing similarities	**sharpening**	noticing differences	Holzman and Gardner, Messic
random	preferring to self-organize materials	**sequential**	preferring materials to be pre-organized	Gregorc
synthetic	assembling pieces into wholes	**analytic**	Disassembl-ing wholes into pieces	Kant

(Leaver, Ehrman and Shekhtman, 2005, p. 71)

Appendix E: Personality Types: How They Like to Learn

Type	How They Like to Learn
ESFJ "extraverted-sensing-feeling-judgers"	cooperative groups
ESTJ "extraverted-sensing-thinking-judgers"	organization, clear instructions, deadlines
ENFJ "extraverted-intuitive-feeling-judgers"	one-on-one or with peer groups
ENTJ "extraverted-intuitive-thinking-judgers"	leading a group of peers in a project
ESFP "extraverted-sensing-feeling-perceivers"	activity with a group and with choice

ESTP "extraverted-sensing-thinking-perceivers"	games, negotiations, simulations
ENFP "extraverted-intuitive-feeling-perceivers"	real-life applications, projects
ENTP "extraverted-intuitive-thinking-perceivers"	analysis, invention, develop new procedure
ISFJ "introverted-sensing-feeling-judgers"	manuals, assisting others
ISTJ "introverted-sensing-thinking-judgers"	details, calculations
INFJ "introverted-sensing-feeling-judgers"	plays, poetry, visual images, archetypes
INTJ "introverted-intuitive-thinking-judgers"	manipulation of theory, logical problems
INFP "introverted-intuitive-feeling-perceivers"	creative writing, metaphor, impressionism
ISFP "introverted-sensing-feeling-perceivers"	practice, play, action, concretization
ISTP "introverted-sensing-thinking-perceiving"	outdoors activities, artwork
INTP "introverted-intuitive-thinking-perceivers"	research, systematize, theorize

(Leaver, 1998, p. 31).

Appendix F: Student Questionnaire

1. What strategies do you employ to learn target language vocabulary?

2. What strategies do you employ to learn target language grammar?

3. What strategies do you use to enhance your overall speaking ability?

4. What strategies do you employ to improve your overall narration performance?

5. What strategies do you employ to improve your past narration performance?

6. What seems to be working well with you now?

7. What does not seem to be working so well?

8. What would you like to change?

(adapted from Ehrman, 1996, p. 38)