AGE AND KNOWLEDGE OF MORPHOSYNTAX IN ENGLISH AS AN ADDITIONAL LANGUAGE: GRAMMATICAL JUDGMENT AND ERROR CORRECTION

By Muhammad Asif Qureshi

A Dissertation

Submitted in Partial Fulfillment

of the Requirements for the Degree of

Doctor of Philosophy

in Applied Linguistics

Northern Arizona University

May 2015

Approved:

Mary McGroarty, Ph.D., Co-chair

Luke Plonsky, Ph.D., Co-chair

Fredricka L. Stoller, Ed.D.

Norbert Francis, Doctor of Pedagogy

UMI Number: 3705456

All rights reserved

INFORMATION TO ALL USERS

The quality of this reproduction is dependent upon the quality of the copy submitted.

In the unlikely event that the author did not send a complete manuscript and there are missing pages, these will be noted. Also, if material had to be removed, a note will indicate the deletion.



UMI 3705456

Published by ProQuest LLC (2015). Copyright in the Dissertation held by the Author.

Microform Edition © ProQuest LLC.
All rights reserved. This work is protected against unauthorized copying under Title 17, United States Code



ProQuest LLC. 789 East Eisenhower Parkway P.O. Box 1346 Ann Arbor, MI 48106 - 1346

ABSTRACT

AGE AND KNOWLEDGE OF MORPHOSYNTAX IN ENGLISH AS AN
ADDITIONAL LANGUAGE: GRAMMATICAL JUDGMENT AND ERROR
CORRECTION

MUHAMMAD ASIF QURESHI

Research on age and second language acquisition (L2A) is vast, but inconclusive. Such research has mainly been motivated by the Critical Period Hypothesis (CPH), which postulates that language acquisition becomes extremely difficult after the onset of puberty. Also, there is a lack of research on age and third/additional language (L3/Ln) learning. To fill this gap, this dissertation examines differences in morphosyntactic knowledge between early and late learners of English as a L3/Ln. In this study, 'early' and 'late' learners are those participants first exposed to English as a medium of instruction (MOI) in 1st and 11th grades, respectively. Participants' morphosyntactic knowledge was assessed based on two tasks: (a) a Grammaticality Judgment Task (GJT) and (b) an editing task, which required participants to correct morphosyntactic errors. Three hundred and thirty five undergraduate and graduate students from two universities in Pakistan voluntarily participated in the research.

Results of the group comparisons showed no statistically significant differences between early and late learners on the GJT; however, on the editing task, a modest but significant difference was observed between the two groups, with late learners scoring higher. This finding contradicts the predictions of the CPH.

On individual morphosyntactic features in the GJT, a significant difference was observed between the two groups on past tense and third person singular. The effect sizes supported an edge for late learners. In contrast to the GJT, on the editing task all morphosyntactic features (a total of eight features) except adverb suffix, present progressive, and past tense showed a small but significant difference, again favoring late learners. In terms of task difficulty, both groups attained higher scores on the GJT and lower scores on the editing task. Also, a strong and statistically significant correlation was found between scores for grammatical and ungrammatical stimuli on the GJT, but a very weak and statistically non-significant correlation between the grammatical and ungrammatical halves of the GJT and the editing task.

Results showed that early L3/Ln learners did not have an edge over late L3/Ln learners in their morphosyntactic proficiency in this English as an additional language context. This dissertation explored L3/Ln learning by predominantly Urdu and Punjabi bilingual speakers, a previously unexplored population. The two measures used provided complementary perspectives on grammatical knowledge. Future research should also examine early and late proficiency differences using a more ecologically valid measure (e.g., a writing task).

Acknowledgments

I would like to express my heartfelt gratitude to my dissertation co-chairs, Dr. Mary McGroarty and Dr. Luke Plonsky, who worked with me through the completion of my dissertation. Without their guidance in piloting and revising instruments, in analyzing data, and in interpreting results, I would not have completed this dissertation successfully. Thank you Dr. McGroarty and Dr. Plonsky. I would also like to express my gratefulness to my committee members, Dr. Fredricka L. Stoller and Dr. Norbert Francis. Dr. Stoller guided in creating thorough scripts for data collection, in creating tables of specifications, and in developing answer keys for data coding. Dr. Francis' help clarified my motivation for the research and the interpretation of its findings. I would also like to thank the Director of the First-Year Writing Program, Dr. Gregory Glau and the Director of the Program in Intensive English, Ms. Jackie Evans for permission to pilot instruments with students enrolled in their respective programs. I consider myself truly lucky to have worked with all of them.

I would also like to extend my gratitude to National University of Modern

Languages (NUML), Islamabad, and the Commission on Science and Technology for

Sustainable Development in the South (COMSATS) for helping me in data collection.

Special thanks go to Dr. Shaheena Bhatti from NUML, and Dr. Sohaila Jvaid and Dr.

Aliya M. Zafar from COMSATS for their help in making data collection possible at their respective institutions. My colleagues Ms. Meghan Moran, Mr. John M. Dunn, and my wife Nosheen Malik helped through different stages of data entering, coding, and cleaning. I truly appreciate their help.

This project was supported by a Doctoral Dissertation Grant awarded by the International Research Foundation (TIRF) for English Language Education and by a Special Project Mini Grant awarded by Arizona Teachers of English to Speakers of Other Languages (AZ-TESOL). Thanks to TIRF and AZ-TESOL for their financial support.

I would like to thank my wife Nosheen, my daughter Nawal, and my son Hamza for their long-term sacrifices (i.e., 5 years) when I couldn't give them their due share of my time. Throughout this dissertation, my entire family adjusted all their routines based on my schedule. This Ph.D. would not have been possible without all their sacrifices and well wishes.

A share of this accomplishment also goes to my friends at NAU. I will always relish the memory of endless arguments, conference discussions, research collaborations, picnics, and parties I had with my friends.

My gratitude also goes to my cousins Mr. Athir Habib and Mrs. Uzma Athir for their help and encouragement during my M.A. and Ph.D. in the US. They were the first ones who hosted me in the US. Their two cute kids, Tamur and Zara, were the joy of my life during my stay with them.

Last, but not the least, this Ph.D. couldn't be possible without the untiring and never wavering support I received from my parents and siblings. Having and raising kids on my own in the U.S. has made me truly appreciate the toil parents could have taken in raising me up. This would not have been possible without Ami (Mom) and Abu's (Dad) prayers and sacrifices. Also, my Ph.D. at NAU would not have been possible without my eldest brother, Bilal Ahmed, who set out the course for the journey that took me to this great accomplishment.

© 2015

By Muhammad Asif Qureshi

Table of Contents

Abstracti
Acknowledgmentsi
Table of Contents vi
List of Tablesi
List of Figures x
I. Chapter 1: Introduction
d. Research Questions 4
III. Chapter 3: Methods 42
A. Number of Participants Needed 42
1. Power Analysis 42
2. Participant Selection 4.
3. Controlled Background Variables 4
B. Instruments 44
1. Background questionnaire 4:
2. Grammaticality judgment task 4:
3. Editing task 50
C. Data Collection Procedures 52
D. Data Preparation and Analyses 53
1. Background questionnaire 53
a. Participants 53
b. Age of exposure (AoE) to second language 54
c. Age at testing (AaT) 54
d. Participants' background information 55
2. Grammaticality judgment task 55

	3. Editing task
Е	Assumption Check for Statistical Analyses
	. Analysis
IV.	Chapter 4: Results
A	
В	8. Question 2a
	C. Question 2b
D	O. Question 3
V C	hapter 5: Discussion and Conclusion
V. C	Summary of Major Findings
A D	B. Implications
Б	1. Theoretical
	2. Methodological
	3. Policy
C	2. Limitations and Directions for Future Research
VI. R	eferences
VII.	Appendices
A	
В	
C	
D	
E	j U
F	\mathcal{F}
G	J &
Н	\mathcal{F}
I.	, E
J.	1
K	1 6
L	
N	\boldsymbol{c}
N	\mathcal{E}
O	$oldsymbol{arepsilon}$
P	. Coding Guide for the Editing Task

List of Tables

1	Mean Differences (d) and Correlations (Zr) for the Impact of Age on L2 (Arranged by Effect Size (d))			
2	Mean Differences (d) and Correlations (Z r) for the Impact of Age on FL			
Proficiency				
3	Mean Differences (d) and Correlations (Zr) for the Impact of Age on L2			
	Proficiency for Timed and Untimed Conditions of Testing			
4	Mean differences (d) and Correlations (Zr) for the Impact of Age on L2 and FL			
	Acquisition and Timed vs. Untimed Conditions			
5	Studies Using Editing Tasks for Error Correction			
6	Frequency of Morphosyntactic Features Investigated in Four Studies			
7	Subcategorization Errors and the Rules They Violate			
8	Grammatical Rule Types and Error Subcategories in Grammaticality Judgment			
•	Test			
9	Participants' Background Information for Early and Late L3/Ln Learners (n = 335)			
10	Language Background of the Participants			
11	The Grammaticality Judgment Task: Item Reliability Scores (KR-20)			
12	The Editing Task: Kappa Reliability Coefficient			
13	The Editing Task: Item Analysis			
14	The Editing Task: Item Reliability (Internal Consistency: KR-20)			
15	Independent Sample t Test for Early and Late L3/Ln Learners on			
	Grammaticality Judgment Task ($n = 335$)			
16	Independent Sample t Test for Early and Late L3/Ln Learners on Editing Task $(n = 311)$			
17	Independent Sample t Test for Early and Late L3/Ln Learners on Individual			
	features on the Grammaticality Judgment Task ($n = 325$)			
18	Percentage Correct for Early and late L3/Ln learners on the Grammaticality			
	Judgment Task			
19	Mann- Whitney U Test for Early and Late L3/Ln Learners on Individual			
	Features on the Editing Task ($n = 311$)			
20	Percentage Correct for Early and Late L3/Ln Learner on the Editing Task			
21	Independent Sample t Test for Grammaticality Judgment Task and the Editing			
	Task (Based on Percentage Correct)			
22	Percentage Correct for Early- and Late L3/Ln Learner on GJT and the Editing			
	Task			

23	EL3Ls' and LL3Ls' Performance on Grammatical Features in the GJT	84
24	EL3Ls' and LL3Ls' Performance on Ungrammatical Features in GJT	84
25	Paired Sample t Test for Grammatical and Ungrammatical Halves of the GJT	
	and the Editing Task (Based on Percentage Correct)	85
26	Intercorrelations, Means, and Standard Deviations for Grammatical and	
	Ungrammatical Halves of the GJT and the Editing Task for Early L3/Ln	
	Learners (n = 225)	87
27	Intercorrelations, Means, and Standard Deviations for Grammatical and	
	Ungrammatical Halves of the GJT and the Editing Task for Late L3/Ln	
	Learners (n = 86)	87
28	Percentage Correct for Early L3/Ln Learners on GJT (Grammatical items) only	
	and the Editing Task	88
29	Percentage Correct for Late L3/Ln Learners on GJT (Ungrammatical items) and	
	the Editing Task	90

List of Figures

1	Normality Plot for the Whole Group on GJT	66
2	Normality Plot for early L3/Ln learners	67
3	Normality Plot for late L3/Ln learners	67
4	Normality Plot for the Whole Group on the Editing Task	68
5	Normality plot for early L3/Ln learners	68
6	Normality plot for late L3/Ln learners	68
7	Percent Correct Per Feature by Early and late L3/Ln learners	76
8	Percentage Correct Per Feature by Early and late L3/Ln learners on Editing	
	Task	79
9	Percentage Correct Per Feature by Early L3/Ln Learners on GJT and Editing	
	Task	81
10	Percentage Correct Per Feature by Late L3/Ln Learners on GJT and Editing	82
	Task	
11	Comparison for Early and late L3/Ln learners Grammatical and	
	Ungrammatical items in GJT with their performance on the Editing Task	91

Chapter 1: Introduction

In an increasingly globalized world, the use and importance of English cannot be over-emphasized. In English speaking countries, its importance is necessitated by the influx of immigrants who arrive there with a hope of finding better opportunities. In non-English speaking countries, a good command of English facilitates, if not assures, achievement of a decent life (Evans & Morrison, 2011; Hamid, 2011). Partly due to the British imperial legacy, and partly due to the advancement of the English speaking world, non-English speaking countries attach great significance to knowing English. For these reasons, several countries use English as their official and academic language and endeavor to introduce such language policies that might help in better teaching and learning of English.

With the need to learn English comes the question when to start learning or studying it. Although several studies have explored the effects of age on language acquisition, results in this area are far from conclusive. For second language (L2) acquisition, the critical period hypothesis (CPH) has been much debated. The CPH, according to Singleton and Munoz (2011), is "a phase in the development of an organism during which a particular capacity or behavior must be acquired if it is to be acquired at all" (p. 408). Although Penfield and Roberts (1959) are credited for first recognizing the superior language learning skills of young children, Lenneberg (1967) popularized this theory for second language learning. Lenneberg believed that the two hemispheres of brain are similar at birth but with maturation lateralization takes place, and during this process the left hemisphere of brain (the dominant hemisphere) become specialized for language learning. This period is called the hypothetical critical period.

Lenneberg divided the CPH in two types: (1) maturational state hypothesis and, (2) exercise hypothesis. According to *maturational state hypothesis* early in life humans have a superior capacity for acquiring languages. This capacity disappears or declines with maturation. Whereas *exercise hypothesis* states that early in life humans have a superior capacity for acquiring languages; if the capacity is not exercised it will disappear or decline with maturation, and if it is exercised, further language learning abilities will remain intact throughout the life (deGroot, 2011; Johnson & Newport, 1989).

Confirming the effects of age, Granena and Long (2013) state that age of significant exposure to L2 explains approximately 30% of variation in ultimate attainment. However, the impact of age has been shown to be different for phonology (Abrahamsson & Hyltenstam, 2009; Bongaerts, 1999; Flege & Liu, 2001), morphosyntax (DeKesyer, Alfi-Shabatay & Ravid, 2010; Granena, 2012), and lexis (Hellman, 2008; Marinova-Todd, 2003). Most previous research confirms strong negative effects of age for L2 phonology, but reports a lack of such negative evidence for L2 vocabulary acquisition. Age effects have been most controversial for morphosyntax.

Statement of the Problem

Supporting the CPH, several studies document a restrictive role of age for L2 morphosyntactic acquisition (e.g., DeKeyser, 2000; Johnson & Newport, 1989; Krashen, Long, & Scarcella, 1979; Patkowski, 1980; Seol, 2005), but other investigations present exceptions (e.g., Bialystok & Hakuta, 1994; Krashen & Harshman, 1972; Snow & Hoefnagel-Hohle, 1978). Results of a meta-analysis (Qureshi, under review) reveal that potential sources of variation in the impact of age on L2 morphosyntactic acquisition might be moderated by at least three factors: (a) the context of learning: second language

(SL) contexts (e.g., Abrahamsson, 2012; Asher & Garcia, 1969; Birdsong & Molis, 2001; DeKeyser, 2000; Johnson & Newport, 1989; Patkowski, 1980; Seol, 2005) or foreign language (FL) contexts (e.g., Asher & Price, 1967; Cenoz, 2002; Muñoz, 2006; 2011; Pérez-Vidal, Torras, & Celaya, 2000), (b) conditions of testing (i.e., timed and untimed), and (c) types of instruments (i.e., GJTs and others). Moreover, age of arrival (AoA) in SL contexts or age of exposure (AoE) in FL contexts, length of exposure (LoE), and age at testing (AaT) appear to be other important moderating variables influencing the outcome of research on age and L2 acquisition.

Age effects have been more prominent for studies that are conducted in SL contexts, in strictly timed conditions, and in studies using some type of GJT. The major moderating variables such as AoE, LoE, and AaT have considerable dissimilarities in SL and FL contexts: most late learners in FL contexts have a comparatively earlier start (e.g., they are exposed to a FL at age 11 as compared to late learners in SL context who get exposed to a SL at age 16 or above), they have a shorter length of exposure, and they also have a younger age at testing. Similarly, variations in the use of instruments (e.g., GJT and written tasks) also vary in SL and FL contexts.

Most studies in FL contexts do not use a GJT. In contrast, in SL contexts, most studies exploring age effects use a GJT. This preponderance of GJTs has been confirmed by a meta-analysis, which found that out of 26 studies exploring age effects for L2 morphosyntactic acquisition, 20 used a GJT. Other instruments investigating age effects have included narrative tasks, global proficiency tasks, and picture description. These tasks are mostly used in FL contexts. Furthermore, features contained in GJTs and other tasks are not similar; hence, conclusions reached cannot be compared.

Another important issue in research on age and language acquisition is its neglect of third language (L3) acquisition. While most previous research explores L2 acquisition in both SL and FL contexts, much of the world is generally multilingual. There are almost 7000 languages spoken in 200 countries (Lewis, 2009, as cited in Cenoz, 2013). Hence, there is a need to explore age effects on L3/Ln and subsequent languages.

Purpose of the Study

This study investigates age effects on English L3/Ln morphosyntactic acquisition. Pakistan, one of the non-English speaking countries that has declared English as the official language along with Urdu (the national language), offers an ideal context for exploring age effects on L3/Ln acquisition.

In Pakistan, the medium of instruction (MOI) in higher education (i.e., post-secondary school) is English. However, in earlier grades – at elementary and secondary school levels – the MOI varies between English and Urdu. Some children from relatively more affluent backgrounds attend private schools where the MOI is English, whereas others who go to public schools receive all instruction in Urdu. However, even in Urdu medium schools, at least one class period is given to teaching English almost every day. Hence, most of the children attending public and private schooling get some exposure to English

Another significant feature in the case of Pakistan is that prior to attending elementary school most students have already acquired two languages. In most cases, these two languages constitute a child's mother tongue and national language (Urdu), respectively. In some cases, when mother and father speak two different languages, children are exposed to three languages that include mother's and father's languages and

the national language. And then, depending on whether children attend an English or Urdu medium school, their age of significant exposure to English varies significantly. Those who attend an English medium school are exposed to English when they are only 5 to 6 years old, whereas those attending an Urdu medium school have significant exposure to English at grade 11 onwards, when English medium instruction becomes compulsory for all science majors and most arts majors. Most students are approximately 16 years of age when this occurs.

The variability in exposure to English (as early as 6, as late as 16) found in the Pakistani setting is comparable to English learning contexts described in most research on age conducted in SL contexts. Hence, unlike findings of many other studies conducted in FL settings, data in the current study reflect participants whose ages of exposure to English are more like the participants in SL contexts.

A further contribution of this study is that it uses an editing task in conjunction with a typical GJT task. The editing task includes nearly all the features found in the GJT used by DeKeyser (2000) and Johnson and Newport (1989), except yes/no questions and wh-questions. All other morphosyntactic features are similar to those included in the GJT used by DeKeyser (2000) and Johnson and Newport (1989). Hence, results of this study document performance on different tasks by early and late learners.

Specifically, this study investigates the following research questions:

- (1) To what extent do early L3/Ln learners (EL3Ls) differ from late L3/Ln learners (LL3Ls) in their
 - a) judgment of grammaticality (GJT), and
 - b) editing/correcting morphosyntactic errors in a written passage?
- (2) To what extent does L3/Ln learners' knowledge vary
 - c) across morphosyntactic features between and within groups, and
 - d) across task types (i.e., GJT and editing task)?
- (3) Are there differences in the way Early and Late learners' scores on the grammatical/ungrammatical halves of the GJT compare to their editing task performance?

List of Important Terms and Abbreviations

This section lists nine major terms and abbreviations that will be used in the study. The terms include second language, foreign language, early and late learners/starters, implicit and explicit knowledge, morphosyntactic features, and error types.

Age of Significant Exposure (AoSE)

Age at which participants were exposed to English language as a medium of instruction.

Early and Late Learners/Starters

Participants in SL contexts are called early and late starters. Participants in FL contexts are called early and late learners. In the dissertation, participants are divided into two groups (i.e., early and late) based on their exposure to English as a medium of

instruction. The early learner group represents those participants who were exposed to English as medium of instruction in grades 1-5 and late learners are those who were exposed to English as a medium of instruction in grades 11 and 12.

Error Types

This term is used to explain the type of violations included for each type of morphosyntactic feature in the grammaticality judgment task (GJT). For example, past tense has three types of violations: (a) past tense marking is omitted, (b) irregular verb is regularized, and (c) regular ending is marked on irregular verbs.

Explicit Knowledge

Learners are aware that they have learned something [morphosyntactic forms] (e.g., N. Ellis, 2008). This also coincides with the part of Krashen's (1976) acquisition-learning hypothesis, which considers language learning as a conscious process that involves attention to language forms and rules.

Foreign Language (FL)

FL is used for learning a different language in contexts where the language is not found outside the classroom.

Implicit Knowledge

[Morphosyntactic] knowledge that is acquired without necessarily being aware of the learning process that might have taken place (e.g., N. Ellis, 2008). Implicit knowledge may also be understood by Krashen's (1976) acquisition-learning hypothesis, which states that language acquisition takes place without conscious attention.

Morphosyntactic Features

A combined term to define morphological and syntactic features investigated in the dissertation study. The thirteen morphosyntactic features included are past tense, plural, third person singular, present progressive, determiner, pronominalization, particle movement, preposition, gerunds, infinitives, auxiliary, word order, and adverb.

Second Language (L2) and (SL)

L2 is used when second language is referred to in a general sense, including both second language (SL) contexts (e.g., where the target language is the first language, for example, English in the US), and foreign language (FL) contexts (e.g., where the target language is not the first language, for example, English in Pakistan).

Third or Additional Language (L3/Ln)

In this dissertation, L3/Ln will be used to denote English as a third, fourth, fifth, or nth number of language.

Abbreviations Used in the Dissertation

Age at Testing (AaT), Age of Arrival (AoA), Background Questionnaire (BQ), Critical Period Hypothesis (CPH), Early Third Language Learners (EL3Ls), Editing Task (ET), English Medium Instruction (EMI), English language program (ELP), First Language (L1), Fundamental Difference Hypothesis (FDH), Grammaticality Judgment Task (GJT), Length of Residence (LoR), Length of Instruction (LoI), Late Third Language Learners (LL3Ls), Medium of Instruction (MOI), Second Language Acquisition (L2A).

Organization of the Dissertation

The first chapter has offered a brief background of the dissertation study and included a statement of the problem, purpose of the study, and important terms used in the study. Chapter 2 presents a review of the literature and offers a theoretical framework for the study. Chapter 3 explains the methods of the study and offers details about participants, instruments, and procedures for data collection and analysis. Chapter 4 presents overall results of the study and the findings for each research question. Chapter 5 concludes with a discussion of major findings, their implications, and directions for the future research.

Chapter 2: Literature Review

Age effects for L2 acquisition are widely researched. This research includes different learning contexts and testing conditions. Moreover, recent research on age has investigated its effects on different types of language knowledge. Despite the long history of research on age and L2 acquisition, some important gaps still remain in the previous inquiries. This chapter, therefore, presents a brief summary of gaps in the previous research. It further describes the educational context in Pakistan to show the relevance and purpose of exploring age effects in those learners who are exposed to English as a MOI at different academic levels. Finally, this chapter presents an overview of past findings of age effects on language acquisition in SL and FL contexts, and in timed and untimed conditions. It also presents findings of a meta-analysis (Qureshi, under review) summarizing the influence of the above-mentioned variables. Moreover, the literature reviewed here summarizes the use of editing tasks for exploring L2 learners' language proficiency with a view to present its suitability for investigating age effects. In order to better account for the age impact on L2 acquisition, the next section summarizes important gaps in the previous research on age and morphosyntax.

Gaps

A review of literature reveals following gaps in research on the impact of age on L2 morphosyntax. First, GJT-based studies primarily depend on intuitive judgments of L2 learners, hence, focus more on implicit knowledge. L2 learners' performance on measures of explicit knowledge, such as an editing task, is not given the same importance.

Second, most of the aurally delivered GJTs are conducted under strictly timed conditions (e.g., Abrahamsson, 2012; DeKeyser, Alfi-Shabatay, & Ravid, 2010; Seol,

2005), which may not be considered optimal as bilingual speakers have been shown to have slower processing ability (McDonald, 2000, 2006). As a result, FL learners may not perform well on the aurally delivered and time-pressured GJTs when compared to untimed tests simply because of their slow processing ability.

Third, most studies conducted in FL contexts separate their participants in under eight and over eleven years of age groups, which is not comparable to the studies conducted in SL contexts that mostly divide groups in under sixteen and above sixteen groups. Hence, for the studies that investigate age effects on L2 morphosyntax acquisition in FL contexts, it is more appropriate to recruit participants in the same or at least similar age groups as are investigated in L2 contexts. Such an approach helps compare the results between the two contexts.

Fourth, it may also help to glean a better picture if age effects are explored on L2 acquisition of such participants whose L1 is not very similar to their L2, for example L1 English and L2 Dutch or L1 Spanish and L2 English. Rather, two very different languages, such as L1 Chinese and L2 English, or L1 Urdu and L2 English, may be more helpful in comprehending age effects.

Along with these methodological issues, research on the impact of age of learning language has largely ignored L3/Ln acquisition. How similar or different are L3-learners from L2-learners in their morphosyntactic proficiency needs to be investigated. Results of a study investigating early and late-L3-learners can be compared and contrasted with the findings of the studies that explored L2 acquisition only.

An approach that might control for the aforesaid methodological weaknesses and address these gaps is to assess L3/Ln learners' use of morphosyntactic knowledge in a

paper-based GJT conducted in a less-rigidly timed condition (e.g., time given to complete test is 5 minutes higher the group average). Such an instrument could be accompanied by an editing task that examines features similar to those focused on in the GJT. In addition, early starters and late starters in the study should have a wider gap in exposure to L3, and their L1 should have a greater typological distance from the target language.

In order to address these issues, the study reported here focuses on the impact of age of learning English as an L3. It specifically investigates L3/Ln acquisition of English morphosyntax in an FL context. Instruments in the study include an editing task along with a written and less-rigidly timed GJT.

However, before going into the literature reviewed for the current dissertation, it is important to understand the educational context in Pakistan. The section that follows provides details of the educational system in Pakistan.

Educational Context in Pakistan

In Pakistan, education is divided into 5 levels from the first grade to the graduate level. From 1st to 12th grade, schooling is divided into three main levels. From 1st to 5th grade, schooling is considered as elementary; from 6th to 10th grade, it is considered as secondary; and 11th and 12th grades are considered as intermediate. After intermediate, students peruse their education for undergraduate and graduate studies.

In Pakistan, the medium of instruction shifts at different levels of public education. Most of the public schools use Urdu as the medium of instruction from 1st to 5th grade. Most private schools, on the other hand, use English as a medium of instruction from 1st to 5th grade. Most public schools offer both elementary and secondary education, hence, in such schools; the medium of instruction continues to remain Urdu from 1st to

10th grade. However, students have an option to switch from an Urdu-medium school to an English-medium school if they move from Urdu to English medium schools. Some students do avail this option by switching from Urdu- to English-medium schools after they finish elementary school and when they start secondary education (i.e., when entering 6th grade). However, when entering the intermediate level, the medium of instruction suddenly changes into English from 11th grade for all science courses, and continues to remain the same for higher degrees (i.e., undergraduate and graduate programs) with the exception of a few courses in arts, which can be pursued in either Urdu or English.

However prior to intermediate, no public school uses English as a MOI; rather only private schools adopt it for instructional purposes. Each public school teaches English as a subject from kindergarten to tenth grade. In certain cases, teaching English as a subject starts in sixth grade. These classes usually meet for 45 minutes for three to five days. If on average, students attend four classes per week, and each class is 45 minutes, they have 180 minutes (3 hours) of exposure to English every week. Multiplied by number of weeks each year (i.e., 40 weeks, excluding summer and winter vacations), they are exposed to English as a subject for 120 hours each year. Schooling from elementary school to tenth grade involves eleven years (including kindergarten); hence, students who go to public schools have roughly 1,320 hours of exposure (i.e., 120 hours per year x total (11) number of years to English as a subject. After tenth grade, all students receive English medium instruction (EMI) in intermediate (i.e., grades 11 and 12) and undergraduate. If on average they receive 5 hours of instruction per day each week, they receive approximately 6,000 hours of instruction in intermediate and

undergraduate. This makes the total number of EMI for LL3Ls to 7,320 (i.e., 1,320 + 6,000).

Those who go to English medium schools from Kindergarten get roughly 6 hours of exposure to English each day for approximately 5 days in a week. This leads to approximately 13,200 hours of exposure (i.e., 1200 hours per year x 11 total number of years) to English as a MOI until the end of the tenth grade. Adding 6,000 extra hours of EMI exposure that they receive in the intermediate and undergraduate, the total number of hours of EMI EL3Ls receive amounts to 19,200 hours.

In terms of quality of instruction, in public schools learners are introduced to basic reading and writing literacies in English, and to some formulaic spoken expressions. Reading starts with the alphabet and leads to learning names of some nouns (i.e., person, places, and things), reading short stories, poems, and essays. Writing, mostly involves reproduction of memorized text in the form of essays, summaries, or interpretation of poems. Speaking is mostly limited to using memorized sentences related to greetings, telling parents' and siblings' names, level of education, and place of residence. Anything beyond this, hardly ever falls in even teachers' own repertoires.

In Pakistan EMI from elementary level is imparted by private schools only. The term 'private schools' might be misleading here. The notion of 'private' is generally understood as 'better'. However, in Pakistani contexts this has exceptions. Although there are some very good English medium private schools, but then there are scores of small private schools open in every other street that offer relatively much lower standard education than what many might consider as 'private education' in the West. However in many cases, their educational quality is better than that offered by public schools.

Considering the above mentioned educational context and MOI shifts in Pakistan, it is important to analyze how exposure to English as a medium of instruction at different levels of education affects later English proficiency. But before exploring the specific questions that this study attempts to answer, it is pertinent to have a detailed review of the literature on age and L2 acquisition of morphosyntax.

Literature Review

Age and SL Acquisition

In naturalistic second language settings, one of the most cited works in support of the CPH for L2 acquisition is Johnson and Newport (1989). Participants included 46 Korean and Chinese speakers whose ages of arrival in the United States ranged from 3 to 39 and who had spent from 3 to 26 years in the US. The participants were tested on a set of 12 rules of morphosyntax, embedded in 276 sentences. Each rule was arranged in a set of 6 to 16 sentences. Johnson and Newport found a correlation (correlation type not mentioned) of -.77 between the age of arrival (AoA) and the acquisition of L2 English morphosyntax. The study concluded that the impact of the critical period is not limited to first language acquisition but extends to second language acquisition as well.

However, Johnson and Newport (1989) highlight several weaknesses in the CPH as proposed by Lenneberg (1967). In the first case, they did not find the complete cessation of the ability to learn a language at puberty. In other words, though post-pubescent learners may not reach levels of proficiency as high as native or early starters, language does not become completely unlearnable for them, a finding supported repeatedly by subsequent studies (DeKeyser, 2000; Seol, 2005). Secondly, Johnson and Newport (1989) also did not find a continuous decline in L2 learners' performance as a

function of AoA; rather they observed noticeable individual differences across the sample.

Building on Johnson and Newport's work, DeKeyser (2000) attempted to assess the impact of the CP on L2 acquisition by evaluating it from the perspective of the Fundamental Difference Hypothesis (FDH). According to the FDH, adults must rely on their problem solving capacity and, in particular on their verbal analytical ability, in order to learn L2 structures. DeKeyser attempted to replicate Johnson and Newport (1989). His subjects were 57 Hungarians, 32 males and 25 females living in the U.S. Forty-two participants (25 males and 17 females) were older than 16 at the time of immigration, while 15 participants (8 females and 7 males) were younger than 16 when they immigrated to the United States. All subjects were tested on a GJT and a language learning aptitude test. The participants also filled out a two-page questionnaire about their education and language background before coming to the United States. DeKeyser found support for the FDH in that an adult could achieve native-like competence but only through high verbal analytical skills and explicit rule learning. This also demonstrated that different structures show different degrees of correlation with the age of arrival. Specifically, it found a low correlation between AoA and word order without an adverb, yes-no questions, and gender related errors. All other structures showed a high correlation with the AoA.

One important difference between Johnson and Newport (1989) and DeKeyser's (2000) study is the reversed pattern of correlation between GJT scores for early and late arrivals. Johnson and Newport found a strong and significant negative correlation (r = -.24) for late arrivals,

whereas DeKeyser found a small and non-significant correlation (r = -.26) for early arrivals and a very small and non-significant correlation (r = -.04) for the late arrivals. In DeKeyser's study, the correlation between age and grammaticality tests was not statistically significant after the data were split for early and late arrivals.

This may be explained with reference to stronger aptitude scores by late learners in DeKeyser (2000). In DeKeyser (2000), later learners had a medium (r = .33) but significant correlation between their scores on the GJT and the Language Learning Aptitude Test (LLAT), whereas for earlier starters, a negligible (r = .07) non-significant correlation was observed. This finding reveals a higher analytical ability by late starters, especially when no significant correlation was observed between the aptitude scores and their AoA. Hence, it might be concluded that higher aptitude by late learners might lead to a mitigating effect on the role of the CPH.

Nonetheless, later studies by DeKeyser (2003a, 2003b, 2006) strengthen the CPH for implicit learning mechanisms for early starters. DeKeyser, Alfi-Shabatay, and Ravid (2010) also found a positive impact of age for early starters (below the age of 18) and of aptitude for young adults (ages 18 to 40). Ellis (2005) too confirms a strong association between (a) age and timed GJTs for early starters and (b) years of formal instruction and untimed GJTs for late starters. Negative correlations between age and ultimate attainment in L2 morphosyntax have been confirmed in many other studies of learners in L2 contexts (e.g., Abrahamsson, 2012; Kim, 1993; Patkowski, 1980; Seol, 2005). Seol (2005) also replicated Johnson and Newport (1989) and supported the CPH for morphosyntactic acquisition.

Other factors have also been shown to affect the relationship between age of acquisition and L2 structural mastery. Typological distance between two languages is one of the important moderating variables. Birdsong and Molis (2001) also replicated Johnson and Newport (1989); their participants were L1 Spanish speakers. Although their study confirmed an overall significant negative correlation (r = -.77) between AoA and GJT scores, it found opposite correlations between age and GJT scores after the participants were split into early and late arrival groups. Similar to DeKeyser (2000), this study found a small and non-significant correlation (r = -.24) between age and GJT for early arrivals. However, for the late arrival group, this study found a strong negative correlation (r = -.69) between age and L2 proficiency.

These findings can be compared and contrasted with Johnson and Newport (1989) and DeKeyser (2000). Birdsong and Molis (2001) contradict the findings of Johnson and Newport (1989) for the early arrival group, but confirm those by DeKeyser (2000). Birdsong and Molis (2001) observed a non-significant and low negative correlation (r = -.24) as opposed to the correlation (r = -.87) found in Johnson and Newport (2000) for the early arrival group. However, their findings confirm DeKeyser (2000) who found a non-significant correlation (r = -.26). However, for the late starters, Birdsong and Molis's (2001) results contradict findings of both the previous studies. Johnson and Newport (1989) and DeKeyser (2000) found a non-significant and negative correlation (r = -.16 and r = -.24), respectively, between age and L2 grammatical proficiency. Conversely, Birdsong and Molis (2001) found a significant negative correlation (r = -.69) between age and L2 grammatical proficiency. This reverse pattern of strong negative correlations for the late starters raises some serious challenges for the CPH theory that postulates that

after puberty (i.e., about age of 15), "age of arrival should no longer predict ultimate attainment" (de Groot, 2011, p. 68).

Moreover, based on their findings, Birdsong and Molis (2001) also claim that fewer years of exposure to English are required of Spanish speakers as compared to Korean or Chinese speakers (see also Bialystok & Miller, 1999). The authors suggest that the smaller typological distances between English and Spanish as compared to Korean or Chinese and English might explain the differential outcome. The issue of typological distance presents an interesting case for the current study which involves Urdu (U) and Punjabi (P) as the participants' first two languages, which are much different from English. Unlike English, both Urdu and Punjabi have SOV word order. Moreover, both are written and read from right to left and mark gender on verbs. Also, Urdu and Punjabi (i.e., Pakistani Punjabi) scripts are derived from a combination of Persian and Arabic scripts which are completely different from English. These two languages thus might present a better picture of age effects on L3/Ln acquisition of English because the typological similarities may not intervene or facilitate L3/Ln acquisition in the same way as they may for participants who speak Spanish as an L1 and attempt to acquire English.

Another study that failed to show support for an early start was conducted by Snow and Hoefnagel-Hohle (1978), who examined the acquisition of Dutch as an L2 by English speakers of different ages. Their subjects were first divided into two groups: beginners and advanced subjects. The beginners were then divided into five age groups in the following way: 3-5 year-olds (n = 10); 6-7 year-olds (n = 8); 8-10 year-olds (n = 10); 12-15 year-olds (n = 9); and 11 adults. The advanced group had 6-7 year-olds (n = 8), 12-15 year-olds (n = 8), and 10 adult participants. The subjects' proficiency was judged

on Wug-tests, sentence completion tasks, sentence translation, sentence judgment tasks, the Peabody Picture Vocabulary Test, story comprehension, and story-telling tasks.

The subjects in the 12-15 age group and the adults made the fastest progress during the first few months of learning Dutch. At the end of the first year, the 8-10 and 12-15 year-olds had achieved the best control of Dutch. The 3-5 year-olds scored lowest on all the tests employed. This study does not support an early advantage for the acquisition of L2 grammatical forms for the early starters.

However, Snow and Hoefnagel-Hohle (1978) examine rates of learning only, and any generalization to ultimate attainment by adult learners should be made with caution. Moreover, the study examines acquisition of Dutch as an L2 by L1 English speakers — two languages that have many typological and orthographic similarities; hence, any generalization to the acquisition of other languages that have greater typological distance may be problematic (Slavoff & Johnson, 1995).

To sum up the research findings of the age impact on L2 acquisition of morphosyntax, it is appropriate to state that the results of the studies are mixed. Some studies show an advantage for early starters (e.g., DeKeyser, 2000; Johnson & Newport, 1989; Seol, 2005), whereas others contradict it (Snow & Hoefnagel-Hohle, 1978). Similarly, AoA does not predict L2 performance for the late starters in DeKeyser (2000) and Johnson and Newport (1989). Both the studies observed a low and non-significant correlation between the AoA and test scores. However, AoA strongly predicts late starters' L2 performance in Birdsong and Molis (2001). In addition, the typological distance between the L1 and L2 (e.g., L1 Spanish and L2 English vs. L1 Chinese and L2 English) might affect L2 acquisition differently.

In order to investigate the causes of inconsistent findings in the previous research and to better understand the impact of age on L2 acquisition, I conducted a meta-analysis of research on age and L2 acquisition of morphosyntax. Details regarding the inclusion and exclusion criteria and the coding scheme are described elsewhere (Qureshi, under review); only the summary of findings is relevant to the literature review section of this dissertation.

In order to meta-analyze previous research on age and ultimate attainment in second language morphosyntax, two types of effect sizes were extracted from the original studies: standardized mean differences (Cohen's d) between different age groups and correlation coefficients (r) of the relationship between AoA and L2 performance. If mean differences and correlation coefficient values were missing, effect sizes were computed by using reported means, standard deviations, sample sizes, t-values, f-values, and/or percentages. If a study compared more than two groups and produced more than one d or r, all the early starting groups (i.e., ≤ 12 or ≤ 15) were compared with all late starting groups (i.e., >12, or >15), and their effect sizes were averaged to obtain a single effect size. For the correlational analysis between age and ultimate attainment, Fisher's Ztransformation was computed to normalize the data (Cooper, 2010; Lipsey & Wilson, 2001). Native language speaker groups were excluded from the analysis. Similarly, studies that compared native speakers with very advanced L2 learners were also excluded from the analysis. The main reason for excluding native speakers was that the principal objective of the meta-analysis was to compare early starters with late starters.

Table 1 provides data from the meta-analysis for the reported and summary effects of age on L2 acquisition. Data in the table are organized according to the *d* value.

Studies with a greater d value than others appear first in the table. Some studies provided two d values. In such cases, smaller d values are provided right after the greater one obtained from the same study.

Table 1

Mean Differences (d) and Correlations (Z_r) for the Impact of Age on L2 (Arranged by Effect Size (d))

Author & year	п	d	r between age of arrival and L2
Bialystok & Miller, 1999	33	3.56	84
Seol, 2005	33	2.83	-1.00
Johnson & Newport, 1991	21	2.46	74
Patkowski,1980	67	2.30	95
Johnson & Newport, 1989	46	2.27	-1.00
Bott, 1993	28	2.50	67
	47	-1.65	n.a
Granena, 2012	100	1.78	.57
DeKeyser et al., 2010	76	1.37	-1.00
	62	1.11	52
Johnson & Newport, 1992	27	.85	60
Shim, 1993	30	74	n.a
McDonald, 2000	28	.69	70
Ball, 1996	100	.49	.23
Bialystok, 1979	317	.24	n.a
Abrahamsson, 2012	220	17	69
DeKeyser, 2000	57	n.a	74
Kim, 1993	60	n.a	.22
Flege et al., 1999	240	n.a	04
Overall	1592	.58 (1.4)	47 (.49)

Note. A negative correlation indicates higher scores for early starters. A positive d value indicates higher scores for lower ages.

Seventeen studies conducted between 1979 to 2012 were included in this meta-analysis. The total number of participants in the meta-analysis was 1,592. Results of the group comparison show that the mean difference between older and early starters in L2 contexts is small to medium (d = .58, SD = 1.4). This difference is small according to

Cohen's (1988) benchmarks as well as those proposed by Oswald and Plonsky (2014) for interpreting effects within second language acquisition in which a mean difference of 0.40 is considered small and 0.70 is considered medium. We can also interpret this result to mean that early starters scored, on average, approximately one-half of a standard deviation above late learners on dependent measures. Correlation analyses provide evidence of a medium to large relationship between AoA and proficiency ($Z_r = -.47$, SD = .49). This correlation indicates that approximately 22% of the variance in participants' scores can be accounted for by the age when they began to learn the target language.

Age and FL Acquisition

Contrary to the general support found for early starters in SL contexts, studies conducted in FL contexts do not always support an early-start advantage. Muñoz (2011) rejects the notion of maturation as a restrictive agent for language learning in an FL context. In her study, early learners (age below 11) did not surpass older learners (age above 11) on three measures of language proficiency (i.e., a global proficiency test, a lexical test, and a phonological identification test). Her participants (n = 141) had spent fourteen years in an instructed-FL context. However, no significant correlation was found between the age of exposure and any of the three measures of proficiency. Her findings are supported by other studies conducted in FL contexts (e.g., Harley & Hart, 1997; Muñoz, 2008).

Similarly, results of Muñoz's (2008) investigations, generally called the Barcelona Age Factor (BAF), do not confirm an early advantage for early learners in an instructed/formal FL context. In the BAF project, the effects of long term (i.e., 8 years) instruction were examined. This project compared learners from four different age groups

(i.e., 8, 11, 14, and 18+) after they had received instruction for 200 hours (short-term), 416 hours (mid-term), and 724 hours (long-term). Early learners did not surpass late learners in the study at any of the testing periods. Muñoz (2011) attributes this finding to the fact that, in FL contexts, early learners do not have the advantage of informal learning as they are not exposed to the target language outside the classroom. Moreover, early learners are not exposed to the target language in FL contexts as extensively as in L2 contexts.

This problem (i.e., the variable quality and quantity of language contact and use) is an important factor under investigation in recent research on the impact of age. Several studies have found that the most successful L2 learners are those who are both formally and informally immersed in the L2 (Moyer, 2004, 2006, 2009; Muñoz & Singleton, 2007). Here it is important to mention that language contact for L2 learners in a FL context includes number of hours, semesters, or years of instruction (Singleton & Muñoz, 2011). Moreover, nature of exposure is usually formal, whereas informal exposure, as is found in an L2 context, is normally non-existent.

The findings of Cenoz (2002) also contradict the notion of an advantage for early learners in an instructed foreign language context. Her study compares the proficiency of 60 students in grades 8 and 11 after six years of exposure to the target language in an instructional context. The mean ages at the time of testing for 8th and 11th graders were 13.1, and 16.2, respectively. The 8th graders were exposed to the target language in the third grade (at 7-8 years old), and the 11th graders were exposed to English in grade six (at 10-11 years old). The participants were assessed on their pronunciation skills, vocabulary, grammar, fluency, and content in oral production, in composition, on a cloze

test, and on a grammar test. In composition, and on the cloze and grammar tests late learners performed better than the early-learners. Although early learners had a higher score for pronunciation, the differences were not statistically significant. Similarly, other studies conducted in instructed/FL contexts do not support an early-advantage for younger groups of learners (e.g., Larson-Hall, 2008; Navés, Torras, & Celaya, 2003; Torras & Celaya, 2001). This lack of support is also reflected in the results of the meta-analysis conducted by the researcher (Qureshi, in preparation) as well. In FL contexts, the early and late learners were considered as those participants who were below and above the age of eleven years because most of the studies in FL-contexts use this age as the cut off age between early and late learners (e.g., Cenoz, 2002; Muñoz, 2011). A summary of the main findings for studies conducted in FL contexts is provided in Table 2. Table 2 displays author and year of the publication of studies, number of participants in each study, and effect sizes (i.e., mean differences and correlations) for the impact of age on FL acquisition.

Table 2

Mean Differences (d) and Correlations (Zx) for the Impact of Age on FL Proficiency

Author & date	n	d	r between age of arrival and L2
Álvarez, 2006	30	2.32	.70
Cenoz, 2002	30	-1.90	12
Harley& Hart, 1997	65	39	n.a
Navés, Torras & Celaya, 2003	55	25	n.a
Muñoz, 2003	139	.16	44
Torras & Celaya, 2001	42	06	n.a
Larson-Hall, 2008	61	02	44
Garcia-Mayo, 2003	104	02	n.a
Overall	526	09 (1.3)	03 (.45)

For FL contexts, eight studies were examined in order to explore the impact of age on FL acquisition. These studies were conducted between 1979 and 2011, and included 526 participants overall. Unlike studies conducted in L2 contexts, which found a mean difference (d = .58) and a moderate correlation (r = -.47), the FL studies found a very small effect size (d = -.09, SD = 1.3) and a negligible correlation (r = -.03, SD = .45). A negative effect size (d) can also be interpreted to mean that the two groups were, on average, only about a tenth of an SD apart from each other. Simply put, they were very similar in their performance on the dependent variable. Hence, unlike the findings in L2 contexts (see Table 1), FL contexts do not support an 'early advantage' for early-learners. Both group comparisons and correlational analysis demonstrate either an advantage for the late learners, or neutral results for early and late-learners.

In addition to age of exposure and contexts of learning, testing conditions (i.e., timed and untimed) have also been proposed to affect L2 performance and the kind of language knowledge that they elicit. Previous research claims that time-pressured GJTs and grammatical sentences within the GJTs measure implicit knowledge of morphosyntax (Bialystok, 1979; Ellis, 1991), whereas ungrammatical sentences and untimed GJTs have been found to measure explicit knowledge of morphosyntax (Ellis et al., 2009). A brief discussion of age with reference to testing conditions is provided in the next section.

Age and L2 Acquisition: Testing Conditions

Timed and untimed testing conditions might differently affect early and late starters. McDonald (2000, 2006) asserts that late starters have lower decoding ability, working memory span, and processing speed; hence, older non-native learners face

problems with aurally (time-pressured) presented L2 sentences. McDonald (2006) confirms this in a carefully designed study. The study reports the impact of processing difficulties on late L2 learners' grammaticality knowledge through two experiments. Experiment 1 examined late L2 learners' working memory, decoding ability, processing speed, and grammatical judgment as compared to native speakers'. Native speakers performed better than the late starters on all the measures. In experiment 2, the researchers placed native speakers under memory-related stressors (e.g., low or high digit load, white noise). In the stressed condition, native speakers performed similarly to late L2 learners. McDonald's findings support a general processing problem for grammatical ability under stressed conditions rather than any specific module age-related deficit for morphosyntactic ability.

Granena (2012) also supports the detrimental effects of timed condition on the performance of L2 learners of Spanish. She probes L2 performance differences on timed and untimed visual and auditory GJTs. The study included 100 participants – 50 early L2 learners (age of exposure 3 to 6 years) and 50 late L2 learners (age of exposure 16 years or greater). The study also contained a control group of native Spanish speakers (n = 20). In general, the study found a significant difference between the native speakers and nonnative early and late starters in their performance, irrespective of modality (i.e., visual vs. auditory) and testing conditions (i.e., timed vs. untimed). However, it was also observed that the non-native speakers performed better on the visual test as compared to the auditory test. An even more interesting finding of the study is that the late starters improved their performance by 20% in the untimed as compared to the timed condition.

This result alludes to the possibility that explicit knowledge and untimed conditions enabled the late starters to make better use of their knowledge.

Table 3 displays the differential effects of timed and untimed conditions on L2 grammaticality judgments meta-analyzed across 20 studies. It displays time span, number of studies, number of participants, and mean differences and correlations for different age groups in both timed and untimed testing conditions.

Table 3

Mean Differences (d) and Correlations (Z_r) for the Impact of Age on L2 Proficiency for Timed and Untimed Conditions of Testing

Time			Timed	(n = 3)	Untimed ¹	(n =22)
1979-2012	k	n	d	r	D	r
	25	2,118	.59 (1.4)	45 (.45)	.54 (.91)	25 (.51)

Note. In timed condition, k = 11 for effect size d, and k = 16 for the correlation analysis. In untimed condition, k = 3 for effect size d, and k = 4 for the correlation analysis.

Overall 25 samples -- 19 for timed and 6 for untimed conditions -- were included in the study. The studies included in the meta-analyses were published between 1979 and 2012 and included 2,118 participants. Results of this study confirm the restrictive role of timed testing conditions on late L2 learners' performance for both the group comparisons and correlation based studies. Nineteen studies that examined L2 learners' performance in timed conditions produced a (d = .59) and a ($Z_r = -.45$). Contrary to this, the studies that assessed L2 learners' proficiency in untimed conditions found a great variability in mean scores. Johnson and Newport (1992) and Larson-Hall (2008) obtained (d = .85) and (d = -.02), respectively. Moreover, the average observed Z_r for the above stated two

¹ There are only three studies in untimed condition and one out of the three, namely Granena (2012), has a very large effect size (d = 1.81). This effect size skews the outcome. The other two studies have a (d = .85 and -.02). If Granena (2012) is excluded, the resulting d is (d = .08).

studies and Granena (2012) was low (Z_r = -25). Here it is important to mention, however, that only three studies were meta-analyzed for the group comparison in the untimed condition and one of the three studies had a very large effect size (d = 1.81).

An overall summary of meta-analytic results from L2 vs. FL and timed vs. untimed conditions is provided in Table 4.

Table 4

Mean differences (d) and Correlations (Z_r) for the Impact of Age on L2 and FL Acquisition and Timed vs. Untimed Conditions

Results	d (SD)	Z _r
Overall	.30 (1.4)	32 (.51)
L2	.58 (1.4)	47 (.49)
FL	09 (1.3)	03 (.45)
Timed	.59 (1.4)	45 (.45)
Untimed	.54 (.91)	25 (.51)

To sum up, the results of studies that examine the impact of age on L2 are mixed. Early learners seem to be outperforming late learners in a naturalistic context, whereas the reverse is observed in FL/instructed contexts. Moreover, processing studies consider age as a moderating variable that leads to lower processing speed rather than suggesting lack of proficiency in a given morphosyntactic ability. This is also confirmed in the results of the meta-analysis reported above, which present a smaller effect size and correlation value for the untimed conditions.

In addition to contexts of learning and testing conditions, types of knowledge that are affected by age (i.e., implicit and explicit) also influence outcomes. Differential results discussed above for the L2 and FL contexts can be better understood if research

on the type of language knowledge is taken into account because "explicit knowledge [unlike implicit language knowledge acquired through natural exposure] can be learned at any age" (Ellis, 2005, p. 150). Hence, early learners may not necessarily surpass late learners in the acquisition of language knowledge when it is formally learned (i.e., through explicit exposure). This is perhaps due to the fact that, in FL contexts, language is learned through formal exposure and early learners do not have the advantage of extensive informal immersion as they do in L2 contexts. Moreover, adult learners, by virtue of their developed cognitive skills, use explicit knowledge as "a tool to achieve control in linguistic problem solving" (Gutiérrez, 2013, p. 23); therefore, they do not lag behind early starters.

In order to better account for younger and late-learners' performance differences, it is important to have instruments that can elicit different types of language knowledge separately. With this objective in mind, the following section offers a selective synthesis of editing tasks used in previous language research, and relevance of an editing task to research on age. This section also presents a review of methodology used in previous research using an editing task, the relevance of an editing task to a grammaticality judgment task, and finally, the potential contribution of an editing task to better understanding different types of L2 knowledge.

Age and Type of L2 Knowledge

Editing tasks: A tool for examining age effects on explicit language knowledge. DeKeyser (2013) considers the distinction between implicit and explicit knowledge as vital for the accurate understanding of age effects on L2 acquisition. Ellis (2005) states that test types (e.g., editing task) that "encourage the use of rules and focus

on form" (p. 162) stimulate explicit knowledge. An editing task generally requires the identification of errors and then provision of correct alternatives, which is use of explicit knowledge.

A brief review of studies that have used an editing task shows that these studies used an error correction task for encouraging interaction among participants (Storch, 2002) or encouraging peer work (Storch, 2007). This task requires error identification and their subsequent correction. In other studies (e.g., Francis, 2012), students were either given options for correction, or they were provided with a table that required participants to fill in the information about the error, its type, and correction (Diab, 2010). In short, these studies encouraged participants to use their explicit knowledge for error correction by using an editing task. For more details about the studies that used an editing task to check learners' morphosyntactic proficiency, see Table 5.

Table 5
Studies Using Editing Tasks for Error Correction

	Research methodology	RQs/Research objectives	Participants	Nature of the text/Item presentation	Errors identification/Categorization	Scoring decisions	Results
Diab (2010)	Experimental design: (a) control group, & (b) experiment group. Classroombased	Does training reduce rule-based and non-rule-based errors through peerediting and self-editing?	n = 40 ESL, freshman students, in an English medium university in Lebanon.	Editing 1st drafts of three argumentative essays on an editing form Identifying 4 types of errors	Rule-based errors (subject/verb agreement, pronoun agreement), Non rule-based (wrong word choice, awkward sentence structure)	Number of errors in second draft of essay two and three in comparison to draft 1 of essay 1	Experimental group (peer-editing) out-performed comparison group on rule-based errors No significant difference on non-rule-based errors
Ferris & Roberts (2001)	Experimental	1. Differences in ss' ability to self-edit across feedback conditions (code, no codes, no feedback) 2. Differences across error type	 n = 46, ESL, two levels below freshman n = 36, ESL, 1-unit grammar for writing. Southeast Asians = 55% 	Ss' own writing for editing Editing task containing 19 error items in the text	Verb errors, noun ending errors, article errors, wrong word, Sentence structure	Correct, Incorrect/no change 99% agreement (interrater)	No differences between the 2 treatment groups (i.e., coded errors, & underlined errors)
Francis (2012)	Classroom- based	To describe tendencies that emerged in inventory of correction & revisions	n = 45, 2nd, 4th,& 6th gradersCountry not reportedPossibly Mexico	Correcting & revising own texts Make any changes that they (ss) considered necessary	(a) Orthographic correction, (b) Morphosyntactic or semantic, (c) punctuation, capitalization, and discourse-level revisions	 Non-effective attempts Effective attempts 	6th graders performed better than 4th graders, and 4th graders better than 2nd graders

Gass (1983)	Classroom- based	What is the function of grammaticality judgment in L2 acquisition?	n = 13, intermediate ESL, & $n = 8$ advanced ESL	Students' self- written and written by other ss. 6 grammatical and 6 ungrammatical sentences Presented randomly	Not mentioned	Correct/incorrect from Standard English point of view	Advanced group was better at actual corrections
Storch (2002)	Classroom- based	Investigates patterns of dyadic interaction among ESL	n = 33 ss enrolled in an ESL program in Australia	Editing task to encourage peer interaction Errors Embedded in text	Past tense, adjective, adverb, third person, plural: number agreement with complement (majority of immigrants)	No scoring	Researchers were interested in dyadic interaction only
Storch (2007)	Experimental	Do learners in pairs perform better than individual learners on an editing task?	n = 66, (4 intactESL classes;highintermediateproficiency	Editing task Editing task containing 19 error items in the text	Errors: in verb tense/aspect; articles (both definite & indefinite); word forms (e.g. using adjectival forms instead of adverbial forms)	Correct/acceptable or incorrect/unacceptable. Inter-rater reliability score was 92%	t-test; no significant difference between paired groups and individual learner's on accuracy scores

Only four studies reported the morphosyntactic features that they investigated.

The morphosyntactic features are grouped in Table 6 by the frequency with which they appeared in the studies.

Table 6

Frequency of Morphosyntactic Features Investigated in Four Studies²

Morphosyntactic features	Number of studies investigating each morphosyntactic feature
Subject/verb agreement, verb forms	4
Articles	3
Wrong word choice	3
Verb tense/aspect, Past tense	2
Article errors	2
Sentence structure	2
Adjectives	2
Adverbs	2
Plurals	1
Pronoun	1
Noun ending errors	1

² Ferris and Roberts (2001), Francis (2012), Storch (2002, 2007)

To be more specific, the grammatical features that the studies in Table 5 focused on were the following: subject/verb agreement, pronoun agreement, and word choice (Francis, 2012); verb errors, noun ending errors, article errors, and wrong word choice (Ferris & Roberts, 2001); past tense, adjective, adverb, third person, and plural: number agreement with complement (Storch, 2002); and verb tense/aspect; articles (both definite & indefinite); word forms (e.g. using adjectival forms instead of adverbial forms) (Storch, 2007). Most of these features are the same as those included in the editing task used in this dissertation study, for example, subject verb agreement, past tense, plurals, articles, and so forth. The similarity between the features used in the dissertation and those mentioned above alludes to the possibility of using the editing task for investigating

age effects on L2 learners' explicit knowledge of morphosyntax. In addition to the feasibility of using an editing task for investigating age effects, it is also important to grasp the methodology used in studies using such a task. In order to accomplish this, the next section will provide a brief overview of methodology reported in studies using an editing task, and its relationship with the dissertation.

Methodological similarities between studies using an editing task and the dissertation research. This part of the synthesis sheds light on the participants included in the studies using an editing task, and the nature of the texts used for editing purposes. It mainly focuses on the error identification process, error categorization, scoring methods, and compares the methodology used in the studies provided in Table 5.

All the studies except Francis (2012) had adult ESL learners as participants. Francis (2012) included 2nd, 4th, and 6th graders as subjects. All the adult ESL participants had intermediate to advance levels of English proficiency. With regard to age, participants in these studies (except Francis, 2012) and the participants in the current dissertation study were roughly the similar age (i.e., post pubescent) at the time of testing.

Only two studies in Table 5 (Storch, 2002, 2007) used a direct editing task, while others used participants' own drafts for editing purposes. Diab (2010) also included an editing form to be filled out with errors corrected in the drafts written by participants.

Like Storch (2002, 2007), the dissertation proposal also contains a separate editing task.

Although used for different purposes, such as observing participants' interactions while editing the text (Storch, 2002), and examining early and late L2 learners' morphosyntactic proficiency, editing tasks in all three studies (i.e., Storch, 2002, 2007; and dissertation research) encourage error identification and their correction. Not only

participants and testing formats have similarities, but error types contained in several studies using editing tasks are also similar.

Error categories were mostly similar in different studies. For example, both Ferris and Roberts (2001) and Storch (2007) included errors related to verbs, articles, and word choice. Storch (2002) included plurals, and Diab (2010) incorporated pronouns in addition to other features mentioned in Ferris and Roberts (2001) and Storch (2007). Francis (2012) had different error categories compared to the other studies. He examined orthographic corrections, morphosyntactic and semantic patterns at the sentence level, and discourse-level revisions with a focus on punctuation and capitalization.

Reviewing these studies was very helpful in understanding scoring and analysis methods for an editing task. Francis (2012) deserves special mention for scoring methods. He divided all correction attempts into two broader categories –non-effective and effective attempts. Non-effective attempts were further divided into four sub-categories:

(a) no change, (b) correct to correct, (c) correct to error, and (d) error to error. None of these attempts led to any improvement in the actual text; hence, all were considered non-effective. The correct to correct category in non-effective attempts affected clarity of meaning in the sentence and made it less coherent, hence was also considered as non-effective. Effective attempts, on the other hand included two sub-categories: (a) 'correct to correct' and (b) 'error to correct'. A correct to correct effective attempt led to some type of improvement in the revised sentence. 'Error to correct' attempts improved the original by making changes that resulted in correct grammatical forms.

The analysis and division of non-effective and effective attempts in Francis (2012) hold great significance for evaluating an editing task. Unlike GJTs, correction in

an editing task is never a fixed phenomenon (i.e., only one correct answer), rather there may be varied but multiple correct options. Although having multiple categories for correctness and incorrectness should be helpful in accurately determining a response, in the current study, all errors were marked only as correct or incorrect, for the sake of a cleaner data set and analysis. In addition to having multiple categories for validating a response, the involvement of more than one rater should help in obtaining a more dependable inter-rater reliability.

Ferris and Roberts (2001), Gass (1983), and Storch (2007) divided participants' responses into either correct/acceptable or incorrect/unacceptable responses. Ferris and Roberts (2001) and Storch (2007) also reported percentage agreement between the raters for determining the grammaticality of a response. No such reliability estimate was provided by Gass (1983).

Two other studies, namely Storch (2002) and Diab (2010), were different from the other studies reported above. Storch (2002) did not analyze participants' corrections because the focus of the study was interaction among the participants, not the ability to edit in itself. Diab (2010), on the other hand, compared participants' errors in the second drafts of essay II and essay III with their errors in essay I. However, details regarding the procedures for determining errors were not reported.

An analysis of the studies reported above has offered insights into the procedures needed for accurately scoring participants' responses on editing tasks. Divisions and subdivisions for identifying the categories of effective and non-effective attempts, as detailed in Francis (2012), are adapted to analyze participants' responses on the editing task in the dissertation research. However, the number of sub-categories for effective and

non-effective attempts is adjusted to match the editing task. For effective attempts, the term 'correct' is used and all non-effective attempts are termed as 'incorrect'. Correct attempts lead from error-to-correct response, and incorrect attempts include unchanged errors and those that lead from error-to-error. In the dissertation, all the target morphosyntactic features under study are provided in their wrong form and their correction is expected from participants. Moreover, to better account for consistency, inter-rater reliability (i.e., Cohen's Kappa) is reported for the correct and incorrect attempts.

Relevance: Editing task and the grammaticality judgment task. Along with age-related constraints on L2 acquisition, recent research has started investigating the type of L2 knowledge that age might influence more. Subsequently, researchers have attempted to ascertain the type of knowledge that different types of instruments might assess. Because GJTs are consistently used in L2 research, it is crucial to investigate the type of knowledge they explore (Douglas, 2001; Purpura, 2004). Opinions on the type of L2 knowledge that GJTs stimulate range from some degree of explicit knowledge (Bialystok, 1979) to probable use of implicit knowledge (Schachter & Yip, 1990) to a combination of the two (Han, 2000; Han & Ellis, 1998). However, GJTs might explore both: implicit and explicit knowledge depending on the features of task design, such as:

(a) time constraints - whether tests are speeded or not (Bialystok, 1979; Ellis, 1991), (b) task stimulus - whether the items investigated are grammatical or ungrammatical (Ellis, 2009; Loewen, 2009), and (c) task modality – whether tests are aural or written (Granena, 2012). Generally, it has been proposed that aural and speeded tasks and grammatical

items on GJTs elicit implicit knowledge while untimed and written tasks and ungrammatical items elicit explicit knowledge.

Ellis (2004) proposes a three-tier process that learners go through while performing a GJT. These include (a) semantic processing, (b) noticing, and (c) reflecting. During semantic processing, L2 learners attempt to understand the meaning of a given sentence; during noticing, learners try to find if a sentence contains any error; and during reflection, they consider what is incorrect in a sentence, and why it is erroneous. Loewen (2009) argues that the first two steps might be equally applicable to both, the grammatical and ungrammatical sentences because in both cases L2 learners need to understand meaning of a sentence and see if there are any errors in the sentence. However for reflection, Loewen (2009) states that it may not happen at all if L2 learners notice a sentence as grammatical because they might stop thinking about the sentence any more. However for ungrammatical sentences, learners need to reflect to find inaccuracies in the sentence; hence, they reflect and use explicit knowledge.

To confirm the relationship between task stimulus (i.e., grammatical and ungrammatical) and type of L2 knowledge (i.e., implicit and explicit), several investigations have been carried on, but their findings remain inconclusive. Bialystok (1979) found that learners used implicit knowledge for both task stimuli, whereas Ellis (2005) and Ellis and Loewen (2007) found that ungrammatical knowledge loaded more strongly on the factor of explicit knowledge in an untimed test.

In addition to the aforementioned task stimulus and type of knowledge relationship, researchers have also attempted to investigate whether early or late learners would perform better on the two types of knowledge. Ellis (2005) and Gutiérrez (2013)

posit that implicit knowledge is limited to only younger learners, but explicit knowledge can be learnt at any age. Moreover, Bley-Vroman's (1988) Fundamental Difference Hypothesis (FDH) posits that adults cannot access an innate mechanism for learning a language implicitly; they must use their problem solving skills (DeKeyser, 2000). Considering the FDH and the theory that explicit knowledge can be learned at any age, it can be hypothesized that late L3/Ln learners would perform better on ungrammatical items than on grammatical items.

One way to check such a hypothesis is to compare younger and late L3/Ln learners' scores on the grammatical and ungrammatical items in a GJT with their performance on another task (Loewen, 2009). Han and Ellis (1998), for instance, used a timed and a delayed GJT, an oral production task, and a meta-language task. The timed-GJT and oral production tasks loaded on a factor of implicit knowledge whereas the delayed-GJT and meta-lingual explanations on an explicit factor.

Following similar procedures, this study used an editing task along with a loosely-timed GJT. It was hypothesized that the grammatical items in the GJT would reflect implicit knowledge, whereas the items in the editing task and the ungrammatical sentences in the GJT would represent explicit knowledge. This assumption is based on research findings that claim that the test types/items that require "analysis" and "conscious awareness" elicit explicit knowledge (Ellis, 2005). It is, therefore, assumed that the ungrammatical items in the GJT will encourage use of analysis for error identification. Similarly, ungrammatical items in the editing task will also require participants to first recognize the error/s and then correct those. This process of error identification and correction requires L2 learners to explicitly understand the error. On

the other hand, grammatical items in the GJT would stimulate a more spontaneous response because the absence of an error does not impose extra demands for identifying an error and then correcting it.

Thus, inclusion of an editing task complements the GJT in a way that provides an opportunity to examine different type of knowledge by comparing participants' responses to the ungrammatical items in the two tasks. Moreover, participants' performance on the two tasks also highlights performance differences of early and late learners in different testing conditions: decontextualized ungrammatical sentences in the GJT and contextualized grammatical errors in the editing task. Hence, one might be considered as requiring more active involvement (i.e., the editing task), and the other, eliciting more passive responses (i.e., the GJT). To sum up, a comparison between participants' performance on the editing task and the GJT helps to illustrate possible age effects for different types of L2 knowledge (i.e., implicit and explicit).

Research Questions

- (1) To what extent do early-L3-learners (EL3Ls) differ from late-L3-learners (LL3Ls) in their
 - a) judgment of grammaticality, and
 - b) editing/correcting morphosyntactic errors in a written passage?
- (2) To what extent does L3/Ln learners' knowledge vary
 - c) across morphosyntactic features between and within groups, and
 - d) across task types (i.e., GJT and editing task)?
- (3) Are there differences in the way Early and Late learners' scores on the grammatical/ungrammatical halves of the GJT compare to their editing task performance?

Chapter 3: Methods

This section is divided into six parts: (a) number of participants needed, (b) instruments, (c) data collection procedures, (d) data preparation and descriptive statistics, (e) a brief report on assumptions check for suitable statistical analyses, and (f) analysis. Each section is sub-divided into subsections and is described in the following sections.

Number of Participants Needed

Power Analysis

To determine the number of participants required for this study, an a priori power analysis (Faul, Erdfelder, Lang, & Buchner, 2007) was run. A power analysis was necessary to determine a sample large enough to correctly reject the null hypothesis that there is no difference between the early L3/Ln group and late L3/Ln group in their performance on the GJT and editing task. The maximum total power of a test is 1.0 and the minimum is zero (0). A power level of .80 is usually considered as acceptable for a particular test in a study (Cohen, 1988; Vogt, 1999). For the current study, a power level of .80 was decided. Two other variables that affect power are effect size and significance (alpha level) (Biostat, 2004). Because the study involved group comparison and correlational designs, the effect size Cohen's d (d = .30) and Fisher's Z-transformation (Z r = - .32) were taken from the meta-analyses on the impact of age on L2 acquisition conducted by the researcher. The alpha was set at .05, which is the standard in most studies in applied linguistics.

The GPower analysis revealed that a sample size (n = 278) for group comparison and (n = 59) for correlational design would be needed to appropriately detect statistical significance.

Participant Selection

To select participants for this study, Institutional Review Board's (IRB) approval was first secured from Northern Arizona University (NAU). To maintain consistency while collecting data from different sites, a script was developed for the consent form. The script and the consent form are in Appendices A and B, respectively. For participant selection, two universities were contacted in Islamabad, Pakistan. Each university was asked to provide a group of at least 200 students to participate in the study. Universities were also requested to provide teachers who could help in data collection. Overall, a group of 409 participants was secured from both universities. After screening and cleaning the data, 335 participants were left for further analysis. At the time of data collection, all participants were undergraduate or graduate students.

Participants consisted of two groups of adults: (a) Early-L3-learners of English and (b) late-L3-learners of English. The EL3L were participants who acquired their first two languages in early childhood through direct exposure to these languages in their homes, but acquired their L3/Ln by attending an English medium school in the elementary grades. Their ages of significant exposure¹ to the L3/Ln might range between four years of age to six years of age. The late L3/Ln learner group consisted of those participants who were exposed to their first two languages (i.e., Urdu and Punjabi) in early childhood, but were not exposed to the L3/Ln (English) as a medium of instruction (MOI) until grade eleven (i.e., approximately 17 years of age). Prior to that, they attended an Urdu medium school where their exposure to English would have been less than 4 to 5 hours per week.

_

¹ Age of significant exposure to L3/Ln is operationalized as the age at which the participants were first exposed to English as a medium of instruction (MOI).

In Pakistan, only private schools use English as a MOI in elementary schools, hence all the participants who were exposed to English as a MOI at elementary schools attended private schooling. Those who started receiving English MOI at eleventh grade, attended public schools. Detailed information about participants' background variables is provided in the analysis section.

Controlled Background Variables

The study aimed to control for two variables: (a) educational level and (b) minimum number of years of significant exposure to the third language (English). At the time of testing, all participants met the following criteria for educational level: (a) they were in the fourth year of an undergraduate program, or attending a master's program and (b) all the participants had at least five years of exposure to English as a MOI. For early-L3-learners, the requirement of being in the fourth year was exempted because they should have had more than 5 years of exposure to English as an L3/Ln by the time they were in college. These requirements established a similar L3/Ln background in terms of minimum length of exposure (MLE) criterion (i.e., 5 years). In order to control for the contextual variability among participants, all the participants studied in the same city, in this case Islamabad, Pakistan.

Instruments

A background questionnaire (BQ) and two instruments to measure morphosyntactic knowledge were included in the study. Also, a consent form to request participants' voluntary participation in the study was obtained from Northern Arizona University. Details for the BQ and each measure are provided below.

Background questionnaire (BQ)

Background information was obtained by administering a background questionnaire. A draft of the BQ was piloted with 10 L2 learners who resembled the actual participants of the study. Based on the pilot, changes were made in the BQ. The final version of the BQ contained 23 questions about participants' educational background, number of languages known, age of exposure to L3, age at testing, and self-rated proficiency in English. Participants' background information obtained through the BQ is detailed in the analyses section. The administration script and BQ are located in Appendices C and D, respectively.

Grammaticality judgment task

A paper-based GJT was used for data collection. The reason for choosing a paper-based GJT as opposed to an aural GJT was motivated by the consideration that participants could solely depend on their morphosyntactic proficiency while attempting the GJT, avoiding the effects of their listening proficiency as an intervening variable. Moreover, previous research shows that the older non-native learners face problems with aurally (time-pressured) presented L2 sentences (e.g., Granena, 2012; McDonald, 2002, 2006). Their lower decoding ability, working memory span, and processing speed interfere with their ability to accurately process sentences (McDonald, 2000). To further control procedural bias that is caused by the speediness of the strictly timed tests, participants were given 25 minutes to complete the task, which was 5.7 minutes higher than the average time (i.e., 19.3 minutes) that the participants spent on completing the task in the pilot.

The GJT examined participants' performance on the 11 morphosyntactic features that are regarded as the most fundamental aspects of English sentence structure (Seol, 2005). The GJT by Johnson and Newport (1989) with modifications made by DeKeyser (2000) was used in the study. Because the study involved two other instruments (i.e., a background questionnaire and an editing task), the GJT was shortened for the present study to minimize fatigue effects.

The GJT in DeKeyser (2000) contained 196 items overall. Each of the 11 morphosyntactic features was further divided based on the rule violation. For each type of rule violation, a cluster of six sentences was provided. Each cluster had correct and incorrect sentences for the rule type that it represented. Some morphosyntactic features, such as third person, determiners, pronominalization, particle movement, and subcategorization had more than six items in some clusters. Four practice items were included as models of the task. For the current study, the GJT was shortened. Steps taken to condense the GJT for the present study are explained in the paragraph that follows.

In the current study, one pair of correct and incorrect sentences was deleted from each rule-type (rule-types are described in Table 8. One item from each rule type was randomly selected and deleted with its pair. Third person, determiners, pronominalization, and particle movement contained eight items for each rule type; hence, to make it similar to other clusters, two pairs were randomly selected and deleted. The cluster of subcategorization was thoroughly examined and it was found that the previous research lacked clear definitions and explanations of this feature. A somewhat limited explanation of this feature was found in Lineberger, Schwartz, and Saffran (1987) and Johnson and Newport (1989). Lineberger, Schwartz, and Saffran (1987) defined

subcategorization in the following words, "the ungrammatical sentences in this section violate lexically stated constraints on the syntactic frames in which given words can occur" (p. 369). Adapting the instrument from Lineberger et al. (1987), Johnson and Newport (1989) provided the following description of the subcategorization, "These items test subjects' knowledge of subcategorization frames of various verbs. In English, individual verbs determine the type of syntactic frames that may follow them. For example, some verbs require a direct object, while others require prepositional phrases. Because the details of these frames are lexically determined, ill-formed sentences could be created by changing the structure of the required frame for a particular verb while keeping the meaning intact" (p.75). However, a close scrutiny of the sentences in this category masked the variability inherent within each pair of sentences. Table 7 presents the sentences with their possible explanation.

Table 7
Subcategorization Errors and the Rules They Violate

	Examples of Subcategorization Errors	Explanation
1.	George says much too softly. George says his prayers much too softly.	Direct object (DO) needed after says
2.	The little boys laughed the clown The little boys laughed at the clown.	Laugh requires a prepositional phrase when used as a transitive verb
3.	John said me that his wife was ill. John told me that his wife was ill.	Wrong word (<i>say</i> means utterance) <i>Tell</i> means communicating information
4.	The student was learning in his room until late last night. The student was studying in his room until late last night.	Wrong word (<i>learn</i> = gain, acquire) Wrong word (<i>study</i> = devote time and attention)
5.	I want you will go to the store now. I want you to go to the store now.	To Infinitive required after want
6.	I hope you to go to the store now. I hope you will go to the store now.	Hope does not require a transitive verb
7.	The man allows his son watch TV. The man allows his son to watch TV.	To Infinitive required after allows
8.	The man lets his son to watch TV. The man lets his son watch TV.	Bare infinitive required after <i>let</i>
9.	The girls want watching TV. The girls want to watch TV.	To Infinitive required after want
10.	The girls enjoy to watch TV. The girls enjoy watching TV.	Gerund required after enjoy

For accuracy of description, only two pairs of infinitives (i.e., numbers 7 and 8 in Table 7) and one pair of gerunds (i.e., number 9 in Table 7) from subcategorization were included in the study. As a result of curtailment in the number of sentences, the resulting items in the GJT are now reduced to 114. A brief description of the morphosyntactic features contained in the GJT is provided in Table 8.

Table 8

Grammatical Rule Types and Error Subcategories in Grammaticality Judgment Test

Morphosyntactic features	Examples of Errors	Rules Violated
Past tense	Last night the old lady die in her sleep. Janie sleeped with her teddy bear last night. A bat flewed into our attic last night.	Past tense marking omitted Irregular verb regularized Regular ending on irregular stem
Plural	Three boy played on the swings in the park. The boy lost two teeths in the fight.Our neighbor bought new furnitures last week.	Plural marking omitted Irregular plural regularized Mass noun used with plural markers
Third-person singular	Every Friday our neighbor wash her car. John can plays the piano very well.	Omission in obligatory context Marked on main verb after modals
Present progressive	Janet is wear the dress I gave her. Tom working in his office right now.	Omission of –ing in obligatory context Omission of auxiliary [be verb])
Determiner	Tom is reading book in the bathtub. The red is a beautiful color.	Determiners omitted Determiners with abstract nouns
Pronominalization	Mike wrote the letter but didn't send The girl cut himself on a piece of glass.	Omission in an obligatory context Gender error
Particle movement	The man climbed the ladder up carefully. Kevin called Nancy for a date up .	Phrasal verb separation not allowed Particle moved too far
Infinitives	The man allows his son watch TV. The man allows his son to watch TV. The man lets his son to watch TV. The man lets his son watch TV.	To Infinitive required after <i>allows</i> Bare infinitive required after <i>let</i>
Gerunds	The girls enjoy to watch TV. The girls enjoy watching TV.	Gerund required after enjoy
Yes-No question	Will be Harry blamed for the accident? Is waiting Sally in the car? Swam Janet in the race yesterday? Where did Arnie hunted last year?	Aux-aux Aux-verb. Aux missing Double tense marking
WH questions	When Sam will fix his car? What they sell at the corner store?	No auxiliary inversion No auxiliary (do) supplied
Word order	The girl the movie likes. Linda a cake baked John. Bites the dog. The students to the movies went. The student eats quickly his meals.	SVDO violated SVIODO violated SV order violated SV PP order violated Adverb-misplacement

Note. There are 4 items for each rule type except gerunds and two categories of infinitives, which had only 2 items, each.

After condensing the GJT, a lexile measure of the sentences contained in the shortened-GJT was calculated. The obtained lexile range was (10L to 270L), which is lower than the lexile range (450L to L275) suggested for 2nd and 3rd graders. Hence, it was assumed that the difficulty level of sentences in the GJT would not interfere with the performance of the participants. All the sentences were randomized in such a way that no two items/sentences representing the same rule type co-occurred. The shortened version of the GJT was piloted with ten L2 learners in a FL context. The instrument showed a reliability coefficient (KR-20) of .81. The attained reliability was considered strong enough to further use the abbreviated version of the GJT for data collection in the actual dissertation

The table of specifications of the GJT, script for the GJT, examples of the GJT, the GJT task for the participants, and the GJT with answers are in Appendices E, F, G, H, and I, respectively.

In order to further explore early and late L3/Ln groups' differences in morphosyntactic proficiency, an editing task was piloted. The subsequent section describes the editing task used in the study.

Editing task

An editing task requires explicit understanding and demonstration of grammatical rules through detection and correction of errors. It can be helpful in better understanding previous claims that late-learners perform better on the measures associated with explicit knowledge, such as recognizing ungrammatical items in a GJT (Ellis et al., 2009), whereas early-learners perform better on measures that check implicit knowledge (e.g., grammatical items in a GJT) (Bialystok, 1979; Ellis, 1991). It

would also be interesting to compare the Early-L3-learners' performance on the editing task with the late learner group's performance. In order to examine the possibility of using an editing task for comparing early and late L3/Ln learners' ability to correct the mistakes in a written passage, a separate instrument was developed.

To develop the editing instrument, a written passage was selected from a Level 5 reading course in a five-level intensive English program. The passage was adapted to match the morphosyntactic features contained in the shortened version of the GJT so that participants' performance on the two tasks could be compared. Hence, two instances of each morphosyntactic error, except subcategorization, contained in DeKeyser (2000) were inserted in the editing task. Instead of subcategorization, two errors related to gerunds and two related to infinitives were inserted. The resulting editing task contained 12 morphosyntactic features. The syntactic features were determiner, pronominalization, particle movement, preposition, gerunds, infinitives, auxiliaries, word order, and adverb; the morphological features included past tense, plural, third person singular, and present progressive. The editing task under discussion contained 2 instances of error for each morphological and syntactic feature. Overall, the editing task contained 24 morphosyntactic violations.

In order to check the reliability and validity of the instrument, it was trialled twice with native and non-native speakers. A Ph.D. student in Applied Linguistics and the researcher coded the editing task. As a result of the pilot analyses, the editing task obtained Kuder-Richardson 20 (KR-20) reliability coefficient of .71, which was considered as an acceptable level of internal consistency. In terms of inter-rater reliability, a higher index of consistency was observed between the two raters. The

obtained Kappa reliability was .91 (p = .000). Both these reliability indexes indicated the suitability of using an editing task as a reliable measure of morphosyntactic proficiency. Details regarding coding of errors are provided in the analyses section.

As a result of the pilot analyses, and for the actual administration of the editing task, several other documents were developed. These included a table of specifications for the editing task (Appendix J), a script for the editing task (K), examples of the editing task (L), the editing task for the participants (M), the editing task with answers in bold (N), and table of potential answers for the editing task (O).

Data Collection Procedures

Two universities situated in Pakistan were contacted for data collection. Both universities agreed to facilitate data collection by providing approximately 200 participants each, a data-collection venue, and teachers who could help during data collection. After all the specific details for data collection were received, dates, venues, and times for data collection were decided.

On the day of data collection, all directions were provided in English because, in Pakistan, the MOI at higher education level is English. During data collection, participants were seated with sufficient space between the desks to prevent copying. Next, all procedures were explained to the participants on a projector. First, participants were provided the informed consent form. The researcher went over the content of the consent form and explained all the important details. While the researcher was reading the consent form, the participants were told to look at the form. After reading the consent form, the participants were requested to ask any questions that they had regarding the

form. Finally, participants were requested to provide their signatures at the bottom of the form.

Once the informed consent process was completed, the participants were directed to look at the first page of the GJT. During this process, all the directions were clearly spelled out and participants were encouraged to ask any questions that they might have about the tasks. Then participants were shown four sample items and their corrections on a projector. A similar process was repeated for the editing task. Finally, participants were asked to fill in the GJT first, then, the editing task, and lastly, the background questionnaire. During data collection, three other teachers assisted in distribution and collection of instruments and in monitoring participants.

Data Preparation and Descriptive Statistics

Background questionnaire

During data cleaning, particular attention was paid to accuracy. The following section describes the steps taken in data cleaning and preparation. It covers three main sections: (a) participants, (b) age at exposure to L2, and (c) age at testing.

Participants

While screening data from the background questionnaire, all the participants who were exposed to English as a MOI at secondary school level (n = 40) were excluded from further analyses. Because this group was relatively much smaller than the other two groups, it was decided to leave their data out of further analyses. Similarly, the participants who were in the late learner group and had an overall less than five years of exposure to English as MOI (n = 12) were also excluded from further analysis. A minimum of five years of exposure to English as MOI was decided as a threshold for

including participants in the study. Also, two participants who did not specify their educational level at exposure to English as a MOI were excluded from the analysis. Those participants who had their entire education in Urdu medium instruction (n = 2) were also excluded. All the participants who were exposed to English as a MOI at intermediate level (grade 11 and onwards) were combined as late learners. This left a sample size of 335 from the initial total of 409. It is important to mention that, after cleaning data from the editing task, more participants were dropped and the total number of participants left was 311. According to the power analysis, a sample size of 278 was needed; the current sample size exceeds this requirement.

Age of exposure to a second language

Information about AoE to a second language was missing for ten of the early-L3-learners. After excluding these participants, the following descriptive statistics were obtained for early L3/Ln groups' AoE: ($mean\ age = 8.57$, SD = 5.79, range 0 - 22). Among the late L3 starters, AoE was missing for four participants. After excluding these participants, the mean age was (M = 7.97, SD = 5.45, range = 0.22).

Age at testing

Like AoE, particular attention was paid to AaT while coding and cleaning data. Thirty two participants' AaT was missing, hence not included in the analysis. After preliminary analyses, early L3 learners' mean age was observed as 22.70 (SD = 2.61, range = 17 - 36). Among the late L3 learners, only eight participants did not report their AaT, whereas one participant's age was quite high (i.e., 47 years). The mean age obtained for the late L3/Ln group was 22.42 (SD = 2.90, Range = 18 - 47). The following section summarizes the major details about participants' background information.

Participants' background information.

Overall 335 participants took part in the study. The early L3 learners had a sample size (n = 225) and the late L3 learners had a sample size (n = 110). All the early learners were exposed to English as MOI at the first grade level and all the late L3 learners had their first significant exposure to English (i.e., all the course materials, instruction, and assessment in English) starting in grade eleven. All the participants reported speaking another language when they were roughly eight and a half years old. Among the early L3 learners, 85 participants were male and 134 were female. Six participants in the early L3/Ln group did not report their gender. Among the late L3 learners, 56 participants were male and 52 were female. Two participants in the late group did not provide their gender. Table 9 presents the major background information about the participants.

Table 9

Participants' Background Information for Early and Late-L3-learners (N = 335)

Characteristics		EL3L $(n = 225)$	LL3L $(n = 110)$
Grade Level of	Exposure to L3	Grade 1	Grade 11
Age at Testing M $Range$ Gender Male Female Not provided Self-reported Proficiency in English (scalable) Listening M (SD) Speaking M (SD)			
	M	22.61 (2.61)	22.42 (2.79)
	Range	17 - 36	18 - 47
Gender			
	Male	85	56
	Female	134	52
	Not provided	06	02
Self-reported Pr	oficiency in English (scale	: 1-5)	
-	• •	2.88 (.85)	2.50 (1.04)
_		2.38 (.91)	2.24 (1.07)
Reading	M(SD)	3.17 (.84)	2.93 (1.07)
Writing	M(SD)	2.69 (.88)	2.60 (1.06)
Educational leve	el		
Undergradu		78	27
Master's	•••	142	82
Not provide	d	5	1
Majors/Facultie			
2	at & Computer Science	68	34
English Stud		51	19
_	nces & Humanities	32	17
Life & Basic		30	26
Not Provide		44	14

According to Table 9, EL3L provided a relatively higher estimate of their self-reported proficiency in listening, speaking, reading, and writing in English. For example, on a 5-point scale, the EL3L reported a mean of 2.38 (SD = .91) for their speaking proficiency LL3L reported a mean of 2.24 (SD = 1.07) for the same skill. However, it is important to note that the late-L3-learners had a greater variability (i.e., greater standard deviation) in their responses to the self-reported proficiency section. At the time of data collection, all the participants were either undergraduate or graduate students. All the participants came

from the four major faculties: Management and Computer Science, English Studies, Social Sciences and Humanities, Life and Basic Sciences. Forty four participants in the early and 14 in the late L3/Ln groups did not report their majors.

Participants were also asked about L1 and L2 spoken at home. Table 10 provides information about the main L1s and L2s, and other languages spoken at home.

Table 10

Language Background of the Participants

Characteristics	Early learners $(n = 225)$	Late learners $(n = 110)$
Main Language Spoken at Home		
Urdu	89	38
Punjabi	10	15
Pashto	5	7
Other Languages Spoken at Home		
English	2	0
Sindhi	0	2
Potohari, Shina, Brushaski	1 (each)	1 (each)
Balti, Hindko,	0	1 (each)
Brushaski	1	0
More than One Language	113	42
Most Common L2 Spoken by 10+ Participants	S	
English	126	55
Urdu	35	25
Punjabi	34	17
Other L2s Spoken at Home		
Persian	5	2
Pashto	4	0
Arabic, Kashmiri, Siriki, Spanish	1 (each)	0
Arabic	0	1
Not Provided	14	6

According to Table 10, most participants in both groups used Urdu, Punjabi, and Pashto as the main second languages spoken at home. Like the language background questionnaire, data were also analyzed and prepared for the GJT and the editing task. The following sections will explain the step taken during data preparation.

Grammaticality judgment task

All the correct items in the GJT were coded 1, and all the incorrect items were coded 0. All the missing values were left as is, without any imputations because this preserves ecological validity of data collection procedures wherein participants do not follow instructions uniformly, skip items because they do not know what to do, or are tired, etc. All those participants whose responses were excluded from the background analyses due to lack of information were also excluded from the GJT analyses. This resulted in an overall sample size of 335: 225 early L3 starters and 110 late L3 learners. Table 11 provides results of the reliability analyses (KR-20) for the GJT as a whole and for each morphosyntactic feature and each error within each morphosyntactic feature separately.

Table 11

The Grammaticality Judgment Task: Item Reliability Scores (KR-20)

Morphosyntactic features	Error Types	Relia	ability
Overall			. 89
Past Tense			.60
	Past tense marking omitted in obligatory context	.27	
	Irregular verbs regularized	.37	
DI I	ending on irregular stem	.27	40
Plural	Dhoral marshing assisted in abligators contact	26	.48
	Plural marking omitted in obligatory context	.26	
	Irregular plurals regularized	.17	
	Mass nouns used with plural marker	.41	
Third Person		5 0	.53
	Third-person -s omitted in obligatory context	.50	
D (D :	Third-person -s marked on main verb after modals	.30	
Present Progressive			.54
	Progressive -ing omitted in obligatory context	.44	
	Progressive auxiliary omitted	.40	
Determiner			.37
	Determiner omitted in obligatory context	.26	
	Determiner used with abstract nouns	.18	
Pronominalization			.30
	Pronoun omitted in obligatory context	.18	
	Gender errors	.12	
Particle movement			.37
	Phrasal verb separation not allowed	.27	
* a	Separation allowed, but particle moved too far	.14	
Infinitive	m 1 C ''	1.6	.14
	To Infinitive required after allows	.16	
0 1	Bare infinitive required after let	.06	
Gerund	Gerund: required after enjoy		.11
Auxiliaries (Y/N)	A [(XXIAA)	1.4	.62
	aux Aux s[(YNAA)	.14	
	aux Verb s[(YNAV)	.38	
	V s[(YNVS)	.29	
WII 0	Double tense marking (YNDT).	.50	
Wh-Questions	N · · · · · (WIDH)		.38
	No aux inversion (WHNI)	.27	
	No aux (WHNA)	.24	
Word Order		_ = =	.66
	S V DO order violated	.52	
	S V IO DO order violated	.26	
	S V order violated	.46	
	S V PP order violated	.22	
	Adverb placement	.03	

Results of the reliability analyses provided weak to moderate reliabilities for individual error types and morphosyntactic features that contained those error types; however, the overall reliability coefficient of .89 was considered strong enough to run subsequent analyses on the entire GJT. Here it is worth mentioning that an increase in the number of items for each error type might improve the reliability score for each error type. After preparing GJT for the analyses, the editing task was examined. Steps taken to clean, code, and analyze data for the editing task are listed below.

Editing task

For the editing task, data preparation involved data cleaning, data coding, and analyzing preliminary data for accuracy and reliabilities. Data preparation started with data coding by two coders. An editing task may elicit different correct responses for the same error, for example,

Sentence with error: They are unable to take any measures for a better future; as a result, many countries **are become poorer**,

First Correct option: They are unable to take any measures for a better future; as a result, many countries **are becoming poorer**,

Second Correct option: They are unable to take any measures for a better future; as a result, many countries **have become poorer**,

Third Correct option: They are unable to take any measures for a better future; as a result, many countries **are poorer**.

Considering potential variability in participants' responses, it was decided that two raters would code approximately 33% (100/409) of responses. The second coder was

a native English speaker enrolled in Ph.D. program in Applied Linguistics. All the correct responses were coded as 1 and all the incorrect and missing responses were coded as 0.

All the responses that were considered correct by both the coders were scored 1.

A detailed inventory of all the possible correct responses for the 24 items included in the editing task is included as Appendix P.

During coding, if some papers fell in consecutive order (i.e., 45, 46, 47), and had exactly the same responses for all the 24 errors, they were considered as potential copiers, and were excluded from further analyses. During this process, 12 papers were excluded out of the 100 papers chosen for coding by the two raters. After coding by the two raters, the editing task obtained a Kappa reliability of .98. Table 12 presents reliability information for the each error type included in the editing task, as well as for the editing task as a whole.

Table 12

The Editing Task: Kappa Reliability Coefficient

Features	Reliability	Significance level		
Overall reliability	.98	<.01		
Past tense	1	<.01		
Plural	1	<.01		
Third person	1	<.01		
Present progressive	.98	<.01		
Determiner	1	<.01		
Pronominalization	1	<.01		
Particle movement	1	<.01		
Gerund	1	<.01		
Infinitives	.96	<.01		
Auxiliaries	.96	<.01		
Word order	1	<.01		
Adverb suffix	1	<.01		

According to Table 12, both raters had perfect agreement on all error types except present progressive, infinitives, and gerunds, and even on these three error types, the disagreement was minute.

After completing the inter-rater analyses, all the data was coded by the principal investigator. During this process, more papers for potential copiers were identified and excluded from the analyses. Overall, 24 such papers were excluded from further analyses. After this omission, 311 participants' papers remained for further analyses. After data coding and cleaning, analyses were run for item difficulty and discrimination. Table 13 displays results of the item analyses.

Table 13

The Editing Task: Item Analysis

	Features	M (Item difficulty: .25 to .75)	Item Discrimination (>.20, >.30)
1	Past Tense 1	.41	.39
2	Past Tense 2	.12	.35
3	Plural 1	.76	.31
4	Plural 2	.08	.20
5	Third Person 1	.65	.35
6	Third Person 2	.17	.40
7	Present Progressive 1	.46	.52
8	Present Progressive 2	.59	.48
9	Determiner 1	.10	.19
10	Determiner 2	.00	.13
11	Pronominalization 1	.26	.20
12	Pronominalization 1	.65	.42
13	Particle Movement 1	.08	.32
14	Particle Movement 2	.41	.47
15	Gerund1	.53	.41
16	Gerund2	.30	.37
17	Infinitive 1	.72	.26
18	Infinitive 2	.53	.56
19	Auxiliary 1	.30	.36
20	Auxiliary 2	.39	.54
21	Word Order 1	.25	.45
22	Word Order 2	.25	.51
23	Adverb Suffix 1	.60	.44
24	Adverb Suffix 2	.65	.40

Item analyses exposed item number 4, 6, 9, 10, and 13 as having low item difficulty scores. However, for item discrimination, only one item (i.e., item 10) appeared to have a low discrimination score. Before excluding any item from further analyses, it was decided to check the overall instrument reliability of the editing task. Table 14 presents reliability coefficient (KR-20) scores for each error type, as well as for the editing task as a whole.

Table 14

The Editing Task: Item Reliability (Internal Consistency: KR-20)

Features	Reliability
Overall reliability	.83
Past Tense	.36
Plural	.13
Third Person	.35
Present Progressive	.71
Determiner	.09
Pronominalization	.10
Particle movement	.27
Gerund	.33
Infinitive	.38
Auxiliaries	.44
Word Order	.48
Adverb suffix	.39

The obtained KR-20 of .83 was considered as reasonably strong to not exclude any item from further analyses. The editing task for 311 participants was kept for further analyses of age effects on early and late L3/Ln learners' proficiency in English morphosyntax. The EL3L included in this analysis totaled 225, and LL3Lwere 86. To examine the differences between early and late L3/Ln learners on the two tasks, assumptions were checked for running an independent-sample *t*-test. The following section details steps taken to check assumptions.

Assumption Check for Statistical Analyses

In order to answer the research questions, independent sample *t*-test analyses were run. An independent sample *t*-test was considered as an appropriate statistical procedure for analyzing group differences on the GJT and editing task between the early- and late L3/Ln groups. An independent sample *t*-test makes three assumptions about the data: (1) two groups have an equal variance on the dependent variable, (2) the dependent variable

is normally distributed, and (3) data for the two groups are independent (Morgan, Leech, Gloeckner, Barrett, 2007). In order to check these assumptions, several steps were taken. Because participants' scores in each group had no effects on the participants' scores in the other group, assumption of independence of data was met.

For checking the assumption of equal variance, Levene's test of equal variance was conducted. The assumption of equal variance is violated if Levene's test of equal variance is significant (Norusis, 1994). However, in this case the assumption of equal variance was met for the GJT and the editing task. For GJT, the actual variances for the early and late L3/Ln learners were 156.67 and 171.38, respectively. The obtained Levene's value was not significant (p = .72). Like GJT, the assumption of equal variance was met for the editing task as well. The attained variances by the early and late L3/Ln learners were 21.24 and 25.08, respectively. The Levene's value was not significant (p = .36).

The assumption of normality of data within each group was checked by examining the histograms for GJT, for the entire group first, and then, for early and late L3/Ln groups separately. The same procedures were run to check the assumptions of normality for the editing task. Moreover, skewness and kurtosis statistics were also computed. For the GJT, histograms for the whole group and then separate histograms for early and late L3/Ln groups are provided in Figures 1, 2, and 3.

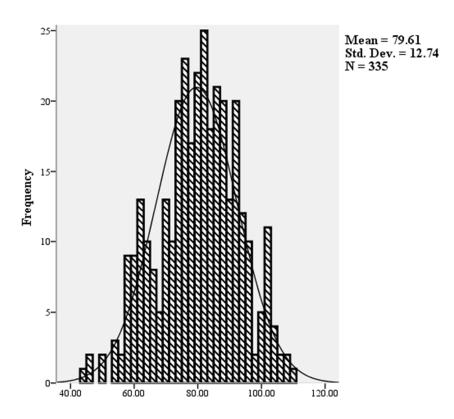
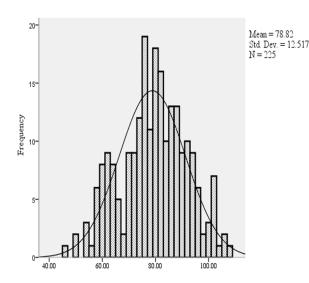


Figure 1: Normality Plot for the Whole Group on GJT

In Figure 1, the Y-axis represents number of participants and x-axis denotes participants' scores on the GJT. The figure shows a bell-shaped distribution of the data which confirms the assumption of normality of data for the early- and late L3/Ln groups combined. According to the figure, data are well-spread between the highest and lowest scores on the task, with an overall mean of 79.61 and standard deviation of 12.74. Skewness and kurtosis values were -1.61 and -1.02, which were less than the Z value of \pm 3.29 (p < .001, two-tailed test) (Tabachnick & Fidell, 2003, 2007); hence, data were considered normal.



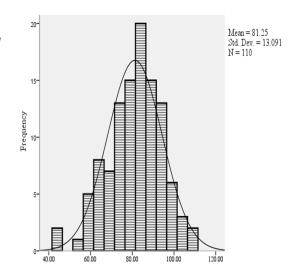


Figure 2: Normality Plot for early L3/Ln learners

Figure 3: Normality Plot for late L3/Ln learners

The histograms (Figures 2 and 3) for early- and late L3/Ln groups also show a normal distribution of data. The data in the two figures represent a bell-shaped curve; hence, they meet the assumption of normality. Skewness and Kurtosis analysis also confirm the assumption of normality of data. Skewness and Kurtosis values obtained for the early L3/Ln group were -0.92 and -1.20, and for the late-L3-learners -1.63 and 0.15, respectively. For both early and late L3/Ln learners, the obtained Z values were lower than the standard Z value of \pm 3.29 (p < .001, two-tailed test) (Tabachnick & Fidell, 2003, 2007), which confirms the assumption of normality of data for the younger and older groups.

As with the GJT, assumptions of normality were checked for the editing task as well. The skewness and kurtosis values for the overall data were 1.42 and -1.77, which were less than the \pm 3.29. Figure 4 provides normality information for the earlier and later L3 groups combined.

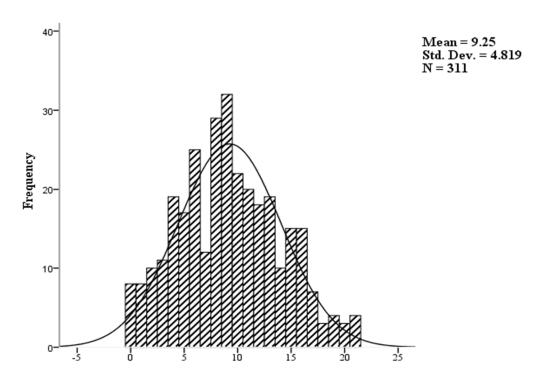


Figure 4: Normality Plot for the Whole Group on the Editing Task

Normality assumptions for the early and late L3/Ln learners were checked separately and histograms for the two are provided as figures 5 and 6.

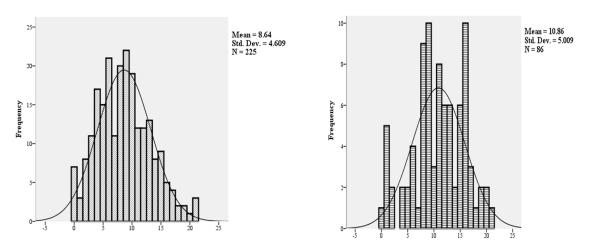


Figure 5: Normality plot for early L3/Ln

Figure 6: Normality plot for late L3/Ln

learners

Visual display of data for the two groups presents clear bell shapes. Skewness and kurtosis analysis also supported the assumptions of normality, and were 2.17 and -0.76

for the early group and -0.99 and -0.86, respectively. Both, the spread of data displayed in histograms, and skewness and kurtosis analyses confirm the assumptions of normality. Hence, it was decided to further analyze the data using an independent sample *t*-test. The next chapter presents results for early and late L3/Ln learners' performance on the two tasks.

Analyses

In order to analyze the three main research questions guiding this study, applicable descriptive statistics, independent and paired sample *t*-tests, and Mann-Whitney U tests were used. For answering question1, parts (a) and (b) that investigate group differences between early late learners on the GJT and the editing task, an independent sample *t*-test was computed. An independent sample *t*-test was chosen after confirming the assumptions of the test, as detailed in the previous section.

Question 2 examined group differences across individual features and across task types. For investigating group differences across individual features on the GJT, an independent sample *t*-test was used, and for examining group differences on the editing task, a Mann-Whitney U test was applied. The Mann-Whitney U test was chosen because the assumption of equality of distribution was violated for all the individual features, and assumption of homogeneity of variance was violated for seven morphosyntactic features in the editing task. To investigate group differences across task type, a paired sample *t*-test was used. Because the two tasks (the GJT and the Editing task) had a different number of items and it was obvious that total maximum number of points for the two tasks would be different, percentages for the number of correct responses were used in

the test. Moreover, percentage correct for EL3Ls' and LL3Ls' were compared on the two tasks across all the morphosyntactic features.

To answer the third research question, an independent sample *t*-test and a paired sample *t*-test was used. An independent sample *t*-test was computed to compare the early and late L3/Ln groups on grammatical and ungrammatical halves of the GJT, and a paired sample *t*-test was used to compare each group's scores on each half of the GJT with their performance on the editing task. To further explore the relationship between the grammatical/ungrammatical stimuli on the GJT and the editing task, a bivariate Pearson correlation was computed. A Pearson correlation was used because both the stimuli on the GJT, and the editing task were normally distributed and assumption of linearity was not violated.

Chapter 4: Results

In the following sections, results are provided for each research question separately. The first research question and its results are presented below.

- (1) To what extent do early L3/Ln learners differ from late L3/Ln learners in their
 - (a) judgment of grammaticality, and
- (b) editing/correcting morphosyntactic errors in a written passage?

 Results of an independent sample *t*-test exposed a small but significant difference between the early and late L3/Ln groups on both tasks, the GJT and the editing task.

 Table 15 displays results of the group comparison for GJT.

Table 15

Independent Sample t Test for Early and Late L3/Ln Learners on the Grammaticality Judgment Task (n = 335)

	n	Mean	SD	t	df	p	d
Early L3/Ln Learners	225	78.81	12.51	-1.642	333	.10	18
Late L3/Ln Learners	110	81.24	13.09				

Note. Total possible maximum points = 114

Results of the group comparison show that the late L3/Ln group scored slightly higher than the early L3/Ln group in overall proficiency on the GJT (t = -1.642, p = .10 > p .05), and the difference between the two groups is small (d = -.18). Both mean scores and negative effect size support better performance by LL3L on the GJT. This finding is interesting because it confirms similar findings in FL contexts (e.g., Cenoz, 2002; Muñoz, 2011) and contradicts those provided for SL contexts (e.g., DeKeyser, 2000; Johnson & Newport, 1989). To further investigate early- and late-learners' differences,

their performance was examined on the editing task. Table 16 provides results of the group comparison for the editing task.

Table 16

Independent Sample t Test for Early and Late L3/Ln Learners on Editing Task (n = 311)

	n	Mean	SD	t	df	p	d
Early L3/Ln Learners	225	8.64	4.60	-3.716	309	.00	46
Late L3/Ln Learners	86	10.86	5.00				

Note. Total possible maximum points = 24

Results of an independent sample t-test again confirmed late learners' advantage on this task whereby they obtained higher mean scores (M = 10.86) than the early-learners who attained a lower mean score (M = 8.64). Differences between the two groups were statistically significant (t = -3.716, p = .00 < p .05, d = -.46), and the effect size was medium and negative, which supports that LL3L performed better on this task as compared EL3L. It is worth mentioning that differences between early and late L3/Ln learners' grammatical knowledge became more obvious on the editing task as compared to the GJT. The next section shows results for different morphosyntactic features and presents results for within and between group analyses.

- (2) To what extent does L3/Ln learners' knowledge vary
 - a. across morphosyntactic features between and within groups, and
 - b. across task types (i.e., the GJT and editing task).
- (a) To explore between group differences, an independent sample *t* test was run. Moreover, EL3Ls' and LL3Ls' mean scores, their standard deviations, and percent correct for each morphosyntactic feature were calculated for the GJT and the editing task. Results for the GJT are reported in Table 17.

Table 17

Independent Sample t Test for Early and Late L3/Ln Learners on Individual Features on the Grammaticality Judgment Task (n = 335)

Morphosyntactic features	EL3L M (SD)	LL3L M (SD)	t	p	d
Past Tense (out of 12)	8.84 (2.17)	9.63 (1.83)	-3.27	.00	39
Plural (out of 12)	8.18 (2.10)	8.44 (2.05)	-1.04	.29	12
Third Person (out of 8)	5.60 (1.60)	6.04 (1.64)	-2.31	.02	27
Present Progressive (out of 8)	6.33 (1.50)	6.54 (1.33)	-1.19	.23	14
Determiner (out of 8)	4.87 (1.60)	4.98 (1.52)	63	.52	07
Pronominalization (out of 8)	5.01 (1.50)	5.08 (1.49)	39	.69	04
Particle movement (out of 8)	5.66 (1.50)	5.52 (1.65)	.79	.42	.08
Infinitive (out of 4)	2.93 (.92)	3.09 (.83)	-1.55	.12	18
Gerund (out of 2)	.78 (.67)	.71 (.73)	.82*	.10	.09
Auxiliaries (Y/N) (out of 16)	11.15 (2.67)	11.13 (2.82)	.06	.95	.00
Wh-Questions (out of 8)	4.58 (1.74)	4.75 (1.45)	95*	.34	10
Word Order (out of 20)	14.93 (2.85)	15.35 (3.06)	-1.21	.22	14

Note. df = 333 for all features except Gerunds (df = 200) and Wh-Questions (df = 255).

^{*} Degree of freedom dropped for these two features because equal variance was not assumed.

According to Table 17, LL3L obtained slightly higher mean scores than EL3Ls on most of the morphosyntactic features except particle movement and gerunds, on which EL3Ls attained marginally higher means. Based on the results of the independent sample t test, the two groups were statistically significantly different on past tense (t = -3.27, p = .00 < p .05, d = -.39) and third person (t = -2.31, p = .00 < p .05, d = -.27), respectively. On both morphosyntactic features, early learners' scores were approximately at the 66th and 62nd percentile of the mean of the late L3/Ln group.

To further explore the EL3L and LL3L differences in grammatical proficiency between the two groups, percentages were computed for the correct responses for each morphosyntactic feature. Table 18 details the group difference between early and late L3/Ln learners.

Table 18

Percentage Correct for Early and late L3/Ln learners on the Grammaticality Judgment Task

Morphosyntactic features	% Correct		
	EL3L	LL3L	
Word Order (out of 20)	93.31	95.93	
Present Progressive (out of 8)	79.12	81.75	
Past Tense (out of 12)	73.66	80.25	
Infinitive (out of 4)	73.25	77.25	
Third Person (out of 8)	70.00	75.50	
Particle movement (out of 8)	70.75	69.00	
Plural (out of 12)	68.16	70.33	
Auxiliaries (Y/N) (out of 16)	69.68	69.65	
Pronominalization (out of 8)	63.00	63.00	
Determiner (out of 8)	60.87	62.25	
Wh-Questions (out of 8)	57.25	59.37	
Gerund (out of 2)	39.00	35.50	

Table 18 complements results reported in Table 17 by showing that the EL3Ls obtained slightly higher percentages than the LL3Ls only on particle movement, and gerunds; on all other morphosyntactic features, however, LL3L learners performed better. If we closely examine the differences in percentages obtained by the two groups, it appears that LL3L attained 5% or higher scores than EL3L on past tense, third person, and

approximately 4% higher score on infinitives. Similarly, on word order, present progressive, plurals, pronominalization, determiners, and wh-questions, late learners obtained 1% to 2% higher scores than early-L3-learners. On the other hand, early-L3-learners achieved higher percentages on plurals, particle movement, and gerunds, by approximately 3, 2, and 4 percent.

Within-group differences on GJT by EL3L and LL3L are visually presented in Figure 7.

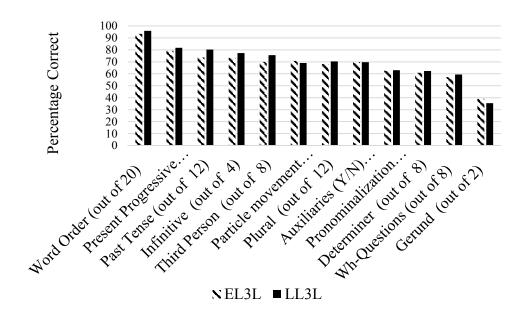


Figure 7: Percent Correct Per Feature by Early and Late L3/Ln Learners
In terms of within group differences across morphosyntactic features on the GJT,
both the groups performed nearly identically, securing highest scores on word order and
lowest on gerunds. Gerunds, wh-questions, determiners, and pronominalizations were the
most difficult features for both groups, and approximately 40% of responses by both
groups were incorrect on these features, except gerunds on which almost 60% responses
were incorrect. However, it is important to mention that there were only 2 items in the
gerund category and results might change with a greater number of items; hence, results

should be interpreted with caution. For auxiliary, particle movement, third person, plural, infinitive, past tense, and present progressive, both the groups attained approximately 70% to 80% percent correct on GJT. Only word order obtained higher than 90% score.

Similar to GJT, EL3Ls' and LL3Ls' performance was measured on the editing task as well. Table 19 provides results of the Mann-Whitney U Test, means and standard deviations for each feature by early- and late-L3 learners.

Table 19
Mann- Whitney U Test for Early and Late L3/Ln Learners on Individual Features on the Editing Task (n = 311).

Morphosyntactic features	EL3L M	LL3L M	U	Z statistics	ap	^b d
Adverb Suffix	1.22 (.84)	1.34 (.69)	8978.00	-1.05	.28	15
Infinitive	1.16 (.73)	1.49 (.71)	7282.50	-3.46	.00	45
Present Progressive	1.01 (.86)	1.13 (.89)	8979.50	-1.04	.28	13
Pronominalization	.85 (.64)	1.05 (.71)	8235.00	-2.252	.02	29
Plural	.78 (.51)	.99 (.54)	7969.50	-2.988	.00	39
Gerund	.76 (.72)	1.05 (.79)	7733.50	-2.936	.00	38
Third Person	.76 (.65)	.95 (.70)	8262.00	-2.19	.02	28
Auxiliaries	.64 (.76)	.80 (.77)	8542.50	-1.74	.08	28
Past Tense	.50 (.62)	.60 (.70)	9028.00	-1.03	.31	15
Word Order	.44 (.66)	.63 (.78)	8513.00	-1.91	.05	26
Particle movement	.43 (.75)	.65 (.66)	7972.00	-2.74	.00	31
Determiner	.08 (.28)	.19 (.39)	8571.00	-2.95	00	32

Note. Total possible maximum points per each morphosyntactic feature = 2

^ap-value is based on exact significance (2-tailed)

 $^{^{}t}bd$ is based on means and SDs.

Results of the Mann-Whitney U Test showed a statistically significant difference between the two groups on all morphosyntactic features except for adverb affixes, present progressive, and past tense. Unlike the results of the GJT where EL3Ls obtained slightly higher scores on at least three grammatical features (i.e., plurals, particle movement, and gerunds); on the editing task, EL3L did not attain higher scores on any of the morphosyntactic features. EL3Ls' and LL3Ls' performance on individual grammatical features was examined by computing percentage of correct responses per each feature. The results for the correct percentage for each morphosyntactic feature are provided in Table 20.

Table 20

Percentage Correct for Early and Late L3/Ln Learner on the Editing Task

Morphosyntactic features	% Correct			
	EL3L	LL3L		
Adverb suffix	61	67		
Infinitive	58	74.50		
Present Progressive	50.50	56.50		
Pronominalization	42.50	52.50		
Plural	39	49.50		
Gerund	38	52.50		
Third Person	38	47.50		
Auxiliaries	32	40		
Past Tense	25	30		
Word Order	22	31.50		
Particle movement	21.50	32.50		
Determiner	4	9.50		

According to Table 20, LL3L attained 10% or higher scores than EL3L on infinitive, pronominalization, plural, gerund, third person, word order, and particle movement; and approximately 5% or higher scores on adverb suffix, progressive, auxiliaries, past tense, and determiners.

For within-group differences across different features, again both groups' performances were nearly identical on the editing task. Within-group differences for EL3L and LL3L are presented visually in Figure 8.

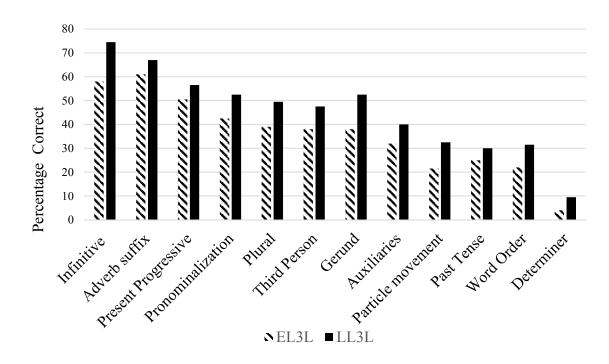


Figure 8: Percentage Correct Per Feature by Early and Late L3/Ln Learners on Editing Task

Both groups obtained the highest scores on infinitives, adverb suffix, present progressive, and pronominalization. However, EL3Ls obtained slightly higher scores on adverb suffix than infinitives whereas LL3Ls attained higher percentages on gerunds than pronominalization, plural, and third person. Particle movement, past tense, word order,

and determiners were relatively more difficult for both groups than the other features on the editing task.

(b) To understand how the two groups might differ in their respective performance on GJT as compared to the editing task, a paired sample *t* test was used. Table 21 displays results of the paired sample *t* test.

Table 21

Paired Sample t Test for Grammaticality Judgment Task and the Editing Task (Based on Percentage Correct)

Tests	n	Mean	SD	t	df	p	d
GJT	335	69.83	11.18	*24.10	477	.01	1.91
Editing Task	311	38.65	20.11				

^{*}Equal variance not assumed

According to Table 21, the two tasks showed a significant difference (t = 24.10, p = .00 < p .05, d = 1.91). The effect size was very large and positive, which means participants achieved higher scores on the GJT as compared to the editing task. To further explore each group's performance on individual features on the two tasks, their percentage corrects were compared. Figure 9 displays EL3Ls' performance on the two tasks.

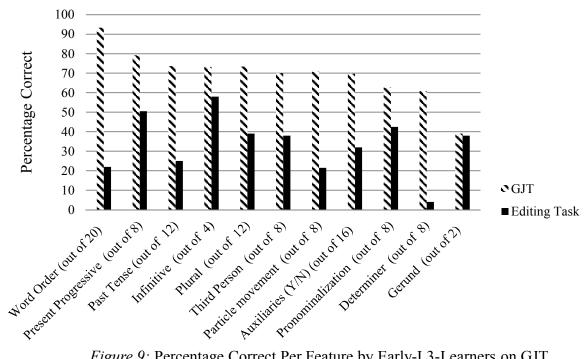


Figure 9: Percentage Correct Per Feature by Early-L3-Learners on GJT and Editing Task

According to Figure 9, a considerable difference in EL3Ls' performance on the two tasks can be seen. On GJT, EL3Ls obtained their highest scores on word order, whereas on the editing task, the same participants attained the second lowest scores on this feature, lower only than determiners. On the editing task, the highest correct percentage were observed for infinitives, whereas on the GJT, present progressive and past tense had higher scores than this morphosyntactic feature. EL3Ls attained lower scores on the GJT than the editing task. The obtained percentages were lower for the following features: word order (70%), determiner (55%), past tense (50%), particle movement (50%), auxiliary (40%), plurals (30%), and third person (30%), respectively. To summarize the EL3Ls' performance differences on the two tasks, it can be said that their performance on the editing task was much lower than the GJT, overall. Also for

EL3Ls, different morphosyntactic features showed different levels of difficulty on the two tasks.

Similar to EL3Ls', LL3Ls also demonstrated considerable gap in their performance between the two tasks. A comparison for late L3/Ln learners' performance on the two tasks is represented visually in Figure 10.

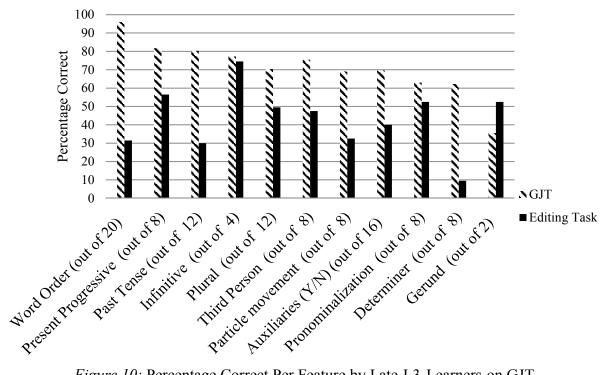


Figure 10: Percentage Correct Per Feature by Late-L3-Learners on GJT and Editing Task

On most of the morphosyntactic features except infinitive and gerund, LL3Ls attained higher percentages on the GJT than the editing task. However, on infinitives, difference in scores between the two tasks was less than 3%, and on gerunds, late L3 learners attained approximately 15% higher scores. Moreover, on the GJT the highest scores were obtained for word order, followed by present progressive, past tense, and infinitive; and lowest scores were attained for gerunds. But on the editing task, LL3Ls achieved highest scores on infinitives, followed by present progressive, pronominalization, gerunds, and

plurals. LL3Ls obtained lowest scores on determiners. A summative comparison of the two groups of learners on two tasks across all morphosyntactic features and within and between group differences is provided in Table 22.

Table 22

Percentage Correct for Early and Late L3/Ln Learner on the GJT and the Editing Task

	I	EL3L		LL3L
	GJT	Editing Task	GJT	Editing Task
Word Order	93.31	22	95.93	31.5
Present Progressive	79.12	50.5	81.75	56.5
Past Tense	73.66	25	80.25	30
Infinitive	73.25	58	77.25	74.5
Plural	73.41	39	70.33	49.5
Third Person	70	38	75.5	47.5
Particle movement	70.75	21.5	69	32.5
Auxiliaries (Y/N)	69.68	32	69.65	40
Pronominalization	62.62	42.5	63.05	52.5
Determiner	60.87	4	62.25	9.5
Gerund	39	38	35.5	52.5

To conclude, EL3Ls' and LL3Ls' ability to judge grammaticality of a sentence in a GJT, and identify and correct an error in an expository text showed remarkable differences within each group and across different grammatical features.

(3) Are there differences in the way early and late learners' scores on the grammatical/ungrammatical halves of the GJT compared to their editing task performance?

To answer question 3, first early and late L3/Ln learners' performance on task stimuli (i.e., grammatical and ungrammatical items) on the GJT was examined. An independent sample –*t*-test did not reveal any significant difference between the early EL3L and LL3L on either one of the two stimuli. Results for the *t*-test analysis are provided in Tables 23 and 24, respectively.

Table 23

Independent Sample t Test for EL3Ls' and LL3Ls' Performance on Grammatical Items in the GJT

	n	Mean	SD	t	df	p	d
Early L3/Ln Learners	225	39.32	6.46	-1.87	333	.06	02
Late L3/Ln Learners	110	40.75	6.86				

Note. Total possible maximum points = 57

Table 24

Independent Sample t Test for EL3Ls' and LL3Ls' Performance on Ungrammatical Items in GJT

	n	Mean	SD	t	df	p	d
Early L3/Ln Learners	225	39.50	6.71	-1.24	333	.21	14
Late L3/Ln Learners	110	40.49	7.00				

Note. Total possible maximum points = 57

Tables 23 and 24 do not show a significant difference between the two groups on the two stimuli; however, there is a very small and negative effect size (d = -.14) for ungrammatical items, which means LL3L obtained slightly higher scores than the EL3L on the ungrammatical stimulus.

To further explore the within-group performance by EL3Ls and LL3Ls, percentages for grammatical and ungrammatical halves of the GJT and the editing task, a paired sample t test was computed. Table 25 provides summary results for a paired sample t test.

Table 25

Paired Sample t Test for Grammatical and Ungrammatical Halves of the GJT and the Editing Task (Based on Percentage Correct)

Morphosyntactic features	n	M	SD	t	df	p	d
Early L3/Ln Learners							
Grammatical	225	68.93	11.30	-6.38	224	.52	04
Ungrammatical	225	69.24	11.76				
Grammatical	225	68.93	11.30	21.93	224	.00	1.46
Editing task	225	36.08	19.21				
Ungrammatical Editing task	225 225	69.24 36.08	11.76 19.21	21.73	224	.00	1.44
Late L3/Ln Learners							
Grammatical	110	71.4	11.99	.60	109	.54	.05
Ungrammatical	110	70.99	12.25				
Grammatical	110	70.44	12.13	10.14	85	.00	1.09
Editing task	86	45.38	20.94				
Ungrammatical	110	70.31	12.01	10.38	85	.00	1.11
Editing task	86	45.34	20.94				

According to Table 25, both learners' groups scored higher on the grammatical and ungrammatical stimuli of the GJT than on the editing task. To further understand the relationship between the two stimuli and the editing task, a bivariate Pearson correlation analysis was computed. Results of the correlations are provided in Tables 26 and 27.

Table 26

Intercorrelations, Means, and Standard Deviations for Grammatical (GJT-G) and Ungrammatical (GJT-U Halves of the GJT and the Editing Task for Early L3/Ln Learners (n = 225) (Based on Percentage Correct)

Tests	GJT-U	EDIT
GJT-G	.80*	01
GJT-U		03
EDIT		

P < .01

Table 27

Intercorrelations, Means, and Standard Deviations for Grammatical and Ungrammatical Halves of the GJT and the Editing Task for Late L3/Ln Learners (n = 86) (Based on Percentage Correct)

Tests	GJT-U	EDIT
1. GJT-G	.75*	.11
2. GJT-U		.17
3. EDIT		

P < .01

Results of the correlation analysis reveal a very strong and positive correlation between the grammatical and ungrammatical halves of the GJT for both early and late L3/Ln groups, r(223) = .80, p < .01, and r(84) = .75, p < .01, respectively. This means learners who had relatively high scores on the grammatical halves of the GJT had high scores on the ungrammatical halves of the GJT as well. However, a similar correlation between the two stimuli on the GJT halves and participants' scores on the editing task was not observed.

To further explore the within-group performance by EL3Ls and LL3Ls, percentages for morphosyntactic features correctly identified for grammatical and ungrammatical items on GJT and their comparison of the similar features on the editing task were computed. Table 28 provides summary results for EL3Ls.

Table 28

Percentage Correct for Early L3/Ln Learners on Grammatical and Ungrammatical GJT Items and the Editing Task

Morphosyntactic features	% Correct EL3L		
	Grammatical GJT	Editing Task	Ungrammatical GJT
Adverb suffix	n.a	61.00	n.a
Infinitive	40.75	58.00	34.75
Present Progressive	84.50	50.50	76.00
Pronominalization	62.00	42.50	60.50
Plural	63.66	39.00	75.66
Gerund	38.00	38.00	37.00
Third Person	75.25	38.00	70.25
Auxiliaries	69.00	32.00	69.00
Past Tense	70.83	25.00	82.33
Word Order	75.6	22.00	75.10
Particle movement	76.5	21.50	61.25
Determiner	59.50	4.00	64.75
Wh-questions	47.00	n.a	45.75

According to Table 28, EL3Ls identified a greater percentage of morphosyntactic features correctly for grammatical stimuli as compared to ungrammatical stimuli. On only three morphosyntactic features, plurals, past tense, and determiners, EL3Ls attained higher scores on ungrammatical items, but on all other features, they obtained higher

scores on grammatical items. In comparison to the editing task, EL3Ls attained lower scores on infinitives and gerunds in ungrammatical stimulus on GJT, and only on infinitives for the grammatical sentences.

To investigate LL3Ls' within-group performance on different stimuli, percentages for morphosyntactic features correctly identified for grammatical and ungrammatical items on GJT and their comparison for the similar features on the editing task were computed. Summary results for LL3Ls' performance on grammatical and ungrammatical items and their comparison with the editing task are provided in Table 29.

Table 29

Percentage Correct for Late L3/Ln Learners on Grammatical and Ungrammatical GJT Items and the Editing Task

Morphosyntactic Features	% Correct EL3L			
	Grammatical GJT	Editing Task	Ungrammatical GJT	
Adverb suffix	n.a	61.00	n.a	
Infinitive	32.50	74.50	34.75	
Present Progressive	74.00	56.50	76.00	
Pronominalization	63.75	52.50	60.50	
Plural	72.66	49.50	75.66	
Gerund	40.00	52.50	37.00	
Third Person	64.75	47.50	70.25	
Auxiliaries	70.63	40.00	69.00	
Past Tense	76.50	30.00	82.33	
Word Order	73.80	31.50	75.10	
Particle movement	65.25	32.50	61.25	
Determiner	62.50	9.50	64.75	
Wh-questions	67.75	n.a	73.25	

Unlike EL3Ls, LL3Ls attained higher percentages on ungrammatical sentences except for pronominalization, gerund, and particle movement. On all other morphosyntactic features in the GJT, late L3/Ln learners obtained higher scores. But in comparison to the editing task, LL3Ls achieved higher scores on the GJT, for both task

stimuli, than on the editing task except for infinitives and gerunds. This might suggest that an editing task is inherently more difficult than a GJT. However, in comparison to EL3Ls, LL3Ls achieved higher scores on both: the ungrammatical stimulus in GJT and the editing task. Figure 11 provides a comparison of early and late L3/Ln learners' performance across two instruments and stimuli.

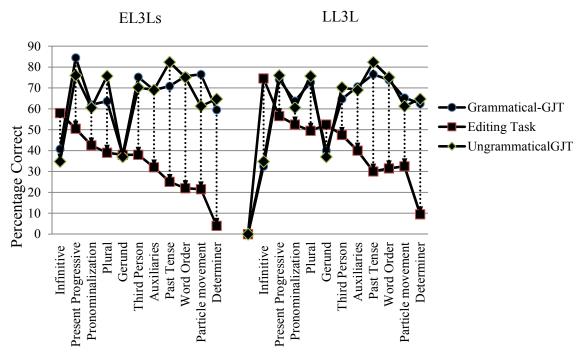


Figure 11: Comparison for Early- and Late-L3-Learners Grammatical and Ungrammatical Items in GJT with Their Performance on the Editing Task

In Figure 11, circles represent grammatical stimulus on the GJT, squares symbolize items on the editing task, and diamonds indicate ungrammatical items on the GJT. As shown in Figure 11, LL3Ls achieved higher percentages on ungrammatical stimuli and the editing task, as compared to EL3Ls. Early L3/Ln learners secured higher percentages on grammatical items, than ungrammatical items in GJT or the editing task within-group.

Chapter 5: Discussion and Conclusion

This study set out to explore the differences in grammatical knowledge between EL3Ls and LL3L in English as an L3. Early L3/Ln learners were first exposed to English as a MOI at the elementary school level, when they were approximately 6 years old. Late L3/Ln learners were first exposed to English as a MOI in grade 11, when they were approximately 16 years old. Participants' knowledge of English grammar was assessed using a GJT and an editing task. Early and late learners' scores were compared between the groups, between the two tasks, and across different morphosyntactic features within each instrument (i.e., the GJT and the editing task). Moreover, participants' scores on grammatical and ungrammatical halves of the GJT were compared with their scores on the editing task.

A total of 411 participants took part in the study. After data cleaning, 335 participants' data were used for analyzing the GJT, and 311 for the editing task. Participants spoke at least two other languages before they were exposed to L3. All the participants were either graduate or undergraduate students and had at least five years of exposure to L3-English as a MOI. Participants' mean ages at testing were 22.61 (SD = 2.61) and 22.42 (SD = 5.64) for the early and late L3/Ln groups, respectively.

The following sections of this chapter provide (a) a summary of major findings in relation to each research question, (b) theoretical, methodological, and language-ineducation policy implications, and (d) limitations and directions for future research.

Summary of Major Findings

Question 1. To what extent do late-L3-learners differ from early-L3-learners in their

- (c) judgment of grammaticality, and
- (d) editing/correcting morphosyntactic errors in a written passage?

To evaluate differences in EL3Ls' and LL3Ls' grammatical knowledge, they were administered a GJT and an editing task. Results of a t-test analysis revealed no significant difference between the two groups on the GJT (t = -1.642, p = .10 > d = -1.8), but on the editing task, the t-value was significant (t = -3.716, p = .00 < p .05). Analysis of mean scores on the editing task revealed that LL3Ls obtained higher scores than EL3Ls, and the effect size was moderate (d = -.46).

Question 2. To what extent does L3/Ln learners' knowledge vary

- (a) across morphosyntactic features between and within groups, and
- (b) across task types (i.e., GJT and editing task)?

Between Group Analysis

A similar though modest late learners' advantage was observed when early and late learners were examined on individual morphosyntactic features between and within groups, as well as across task types (i.e., GJT and editing task). For between groups, late learners attained slightly higher scores on all the morphosyntactic features in GJT except gerunds and particle movement, on which younger learners obtained marginally higher scores.

EL3Ls' performance on two aforesaid features can be compared with DeKesyer (2000) and Johnson and Newport (1989). In both these studies, participants' scores on determiners and subcategorization strongly correlated with age of arrival, which means

the younger a learner is at the time of exposure to the target language, the better they perform on these tasks. Conversely, in the current study, LL3L attained higher means than the EL3Ls on determiners (M = 4.98) and subcategorization (M = 3.82), respectively. For subcategorization, infinitives and gerunds were used in the current study. Also, DeKeyser (2000) and Johnson and Newport (1989) found a strong correlation between participants' ages and their performance on plurals, but the present study does not support this finding. In addition, DeKeyser (2000) did not find age effects for particle movement, but results of this study showed a significant difference between the two groups on particle movement. However, difference in the two groups mean scores was meager -- early L3/Ln group obtained very slightly higher mean scores (M = 5.66, SD = 1.50), as compared to late L3/Ln group that attained lower average scores (M = 5.52, SD = 1.65).

On the editing task, late learners attained higher scores on all the features included in the task. On infinitive, pronominalization, plural, gerund, third person, word order, and particle movement, LL3L obtained approximately 10% higher scores.

Within Group Analysis

For within group analyses, both groups obtained approximately 70 to 80 % correct scores on auxiliary, particle movement, third person, plural, infinitive, past tense, and present progressive on GJT. Wh-questions, determiners, and pronominalizations appeared to be more difficult for both the groups, and approximately 40% of their responses were erroneous on these features. Gerunds appeared to be the most difficult feature, and almost 60% of responses by both groups were incorrect on this feature. In

contrast, word order seemed to be the easiest feature; 90% of responses by both groups were correct on this

On the editing task, adverb suffixes and infinitives appeared to be the easiest tasks and both groups attained scores approximately 60% correct or higher on these features. LL3Ls obtained score approximately 50% correct or higher on seven morphosyntactic features that included adverb suffix, gerund, infinitive, plural, present progressive, pronominalization, and third person. EL3Ls on the other hand, attained scores 50% correct or higher only on three grammatical features: adverb suffix, infinitive, and present progressive; on all other features, their scores ranged between 20% and 40% correct.

To summarize the findings for the between and within group analyses, LL3Ls achieved higher scores on most individual features, in general, and on the editing task, in particular. For within group analysis, both early and late learners showed quite similar patterns of difficulty on features contained in the two tasks, although the easier and more difficult grammatical features were not the same on the two tasks.

Difficulty across Task Types

In terms of difficulty on the two tasks, both groups showed a similar pattern. The GJT appeared to be much easier for both EL3Ls and LL3Ls and their 69.13% and 71.23% responses were correct on it. In contrast, on the editing task, both groups performed more poorly, with only 36% and 45% percent correct. Moreover, each group also differed in performance on the same features (e.g., word order) across the two tasks.

Question 3. Are there differences in the way early and late learners' scores on the grammatical/ungrammatical halves of the GJT compare to their editing task performance?

On the GJT, EL3Ls attained higher scores on all grammatical features except plurals, past tense, and determiners, on which they had higher scores on ungrammatical items. In comparison to their scores on the editing task, EL3Ls obtained higher scores on all grammatical items except infinitives, and on all ungrammatical items in the GJT, except infinitives and gerunds, on which they achieved higher scores on the editing task.

In contrast, LL3Ls performed obtained higher scores on all the ungrammatical features in the GJT except gerunds, particle movement, and pronominalization, on which they obtained higher scores on grammatical items. On the editing task, LL3Ls achieved higher scores only on gerunds and infinitives; on all other features, they had higher correct percentage on the grammatical and ungrammatical features in the GJT.

Implications

The findings of this study have theoretical, methodological, and policy-related implications for research on age and L2A.

Theoretical Implications

For the nature of end-state proficiency in L2A among post-pubescent learners, Birdsong (2009) discuses three theories: first, that views ultimate attainment as impossible; therefore, a failure on the part of learners, second, that considers variability in outcome as a possible option; hence, some learners can attain higher proficiency on some morphosyntactic features, and third, that explores the upper limits of L2A. The current study was based on the hypothesis that some late learners would perform on par with early learners on some morphosyntactic feature contained in the two tasks. This assumption was motivated by the second theory, which postulates that some learners can attain higher scores on some features. Moreover as Muñoz (2011) rejects maturation as

constraint on FL learning, it was assumed that early learners would not outperform late learners. Confirming the theoretical assumption, this research makes an original contribution to SLA in three ways.

First, findings of this study show that the effects of CPH are mediated by the context of learning. Results reinforce previous findings in FL contexts – early learners do not outperform late learners (e.g., Cenoz, 2002; Harley & Hart, 1997; Larson-Hall, 2008; Muñoz 2008, 20011; Navés, Torras, & Celaya, 2003; Torras & Celaya, 2001); rather learners who are exposed to a target language in a later age, obtained higher scores on the GJT and the editing task. The difference between the two groups was significant on the editing task. This finding contradicts a generally held believe of 'earlier is better' in Pakistani FL context, and emphasizes the context-specific reality of second language learning. In target-language deficient context like Pakistan where quality of EMI might be poor, and there is lack of target language use outside the schools context, early starters may not show patterns of proficiency similar to those observed in early acquirers in SL contexts.

Like the nature of early and late proficiency differences, researchers have also argued about the nature of L2 knowledge as static or dynamic (e.g., Larsen-Freeman, 2005). Most studies that use a GJT and no other tasks assume the possibility of only one correct form, which is theoretically and empirically wrong because one and the same error might be corrected in several ways. To account for this theoretical factor, the current study used an editing task that contained morphosyntactic features similar to those contained in other influential studies on age-related research (e.g., DeKeyser, 2000; Johnson & Newport, 1989). Unlike a GJT, an editing task provides opportunity to correct

the same error in multiple ways. By using an editing task, this study explored the dynamic view of language that deems language as a growing and changing phenomenon and expects a multiplicity of correct responses.

Third, results of this study support the Fundamental Difference Hypothesis by Bley-Vroman (1988). According to FDH, adult learners -- who have generally lost the implicit facility for language learning that younger learners have -- mostly depend on their problem solving skills, and for younger learners, this ability is not a strong predictor of language acquisition. In the current study, LL3Ls obtained higher scores than EL3Ls on both the tasks, more specifically on the editing task and ungrammatical items on the GJT, which might be attributed to the nature of the tasks. An editing task and ungrammatical items require more explicit knowledge and problem-solving skills than grammatical items in a GJT. Hence late learners, who have been reported to depend on their problem-solving abilities, showed a clear edge over early learners on the editing task and ungrammatical items in the GJT.

Methodological Implications

Methodologically speaking, age effects in learning English have most often been investigated for learners who come from structurally similar native languages, such as Dutch or Spanish. It is thus important to investigate possible age effects when L3/Ln learners come from native language backgrounds more dissimilar to English, such as Urdu and Punjabi. By investigating age effects on speakers of these two languages, this study advances the scope of research on age and L2 acquisition.

Furthermore, in the previous studies conducted in FL contexts, the gap between early and late starters' age of exposure to target language was much smaller, for example

<11 and >11 in Muñoz (2011), <8 and >11 in Navés, Torras, & Celaya (2003) and Torras & Celaya (2001), and 9 and 12 or 13 in Larson-Hall (2008). In such contexts, learners in both groups might be considered within the same age range, at least in terms of maturation. Most previous studies conducted in SL contexts find a more pronounced difference between early and late learners' scores when groups are split up between <15 and >15 (or >16 or >17) groups (e.g., Birdsong & Molis, 2001; DeKeyser, 2000, Johnson & Newport, 1989). Confirming the practices in SL contexts, the current study included earlier learners who were exposed to L3-English as MOI at the elementary level (1st grade), and later learners who were exposed to L3-English as MOI in intermediate (11th grade in Pakistan). In other words, early and late learners were those who were exposed to L3/Ln before and after the age of 15. Moreover, the minimum gap in the age of exposure to L3/Ln between EL3Ls and LL3Ls was 10 years (i.e., the gap between 1st and 11th grades); hence, the two groups did not fall within the same maturation bracket. By having an age gap between early and late L3/Ln learners similar to those found in studies conducted in SL contexts, this study presents a design that is a methodologically more comparable to studies in SL contexts. Despite this similarity, the results of the GJT in the current study do not confirm findings reported in SL contexts, where early learners outperform late learners; rather, contrary to this, the findings of present study support outcomes reported in FL contexts where no significant difference is observed between early and late learners in their grammatical ability.

Also, this study attempted to validate the results of the GJT. Claims of age have most often been made based solely on results of GJTs, which assess only the ability to recognize correct grammatical forms; it is rare for studies to include any other

instruments to validate GJT results, but the present study did so through the editing task. The findings for the two instruments were quite different (e.g., very low correlation), which implies that GJTs are insufficient as the sole or even primary means of understanding grammatical knowledge. Future studies, therefore, should use more and other measures in addition to GJTs.

Moreover, coding for multiple corrections for the same error on the editing task represent an ecologically more valid way to assess L2 learners' grammatical proficiency. As in reality, language learners might show a range of abilities to correctly use the same grammatical feature in different contexts; thus the editing task used in the study offered a similar opportunity to participants to demonstrate their ability to identify and correct errors in several ways.

Another contribution of this study is the active nature of an editing task. Error identification and correction require active participation by learners. For each correction provided, it can be assumed that a learner thoroughly read the text, identified an error, and then used one of the several options for correction. Such active participation resembles dynamic nature of language learning process, which is different from a relatively more inactive nature of grammaticality judgment that does not involve similar processes.

The order of administration of instruments during data collection might be considered as another contribution of this study. Most studies do not report the order of administration of instruments; however in the methods sections, background information always precedes other instruments, which implies that participants' demographic information was collected before other measures were administered. In a study that

involves a fairly long battery of tests, giving background questionnaire in the beginning might increase cognitive load and lead to fatigue in participants. In the current study, on the other hand, background questionnaire was administered after participants had completed the GJT and the Editing tasks. This was deliberately done so that the learners could focus all their energies on completing the major tasks first, and then fill in the personal information, which might be even easier retrieve. Other studies, I think, would be wise to do the same.

Policy Implications

Because this dissertation examines effects of exposure to English as a MOI at two different academic levels (i.e., elementary and high schools), findings of the study inform language-in-education policies in Pakistan, and perhaps elsewhere, with similar contexts. The early learners in FL contexts do not have the language immersion opportunities similar to those that they might have in SL contexts. Thus, they may not fully benefit from the latent mechanisms that help them in language acquisition. Conversely, later learners by virtue of their problem-solving ability might better learn grammatical aspects of language than earlier learners in FL contexts. Findings of the current study reflect outcomes similar to those in FL contexts. Early learners in the current study did not achieve higher scores on either the GJT or editing task. This was despite the fact that they were exposed to English medium instruction for approximately 19,200 hours as compared to late L3/Ln learners who received English instruction for roughly 7,320 hours only. This outcome is consistent with findings of other studies in FL contexts where early learners did not outperform late learners even when they were exposed to the target language for various lengths of time, for example 200, 416, and 720 hours in

Álvarez (2006), 564 hours in Cenoz (2002), 4.5 years in Larson-Hall, and 2400 hours in Muñoz (2011).

There is a widely held belief that 'younger is better', which is true for SL contexts, where SL learners are exposed to a target language both formally and informally (i.e., in society). But preponderance of evidence suggests the case to be different for FL contexts (e.g., Cenoz 2002, Muñoz, 2006; 2011), where a similar exposure to target language is missing outside the classroom. Hence, in FL contexts, introducing FL programs at an earlier age is a misapplication of L2 research (Spada, 2015).

This study looked only at receptive grammatical knowledge as shown in a GJT and at a semi-direct control of grammar shown on an editing task. There was generally no significant difference between the groups, with relatively small significant differences on some features that favored the later MOI group. As an implication of this finding, governments that plan to introduce a foreign language as a MOI might consider practical reality of their contexts – late starters can attain higher scores on grammaticality judgment and editing task.

However, this does not mean it's a total waste to start English MOI earlier. Other areas of language proficiency, such as pronunciation, reading and listening comprehension, grip on vocabulary depth and width, understanding and use of L2 collocations, and writing ability might be other important dimensions of L2 proficiency, but are not covered in the current study. Proficiency in a second language pronunciation is another very important dimension of age related research, however, it may not be applicable to FL contexts, hence, may not be implicated with regard to language policy.

Limitations and Directions for Future Research

This study makes several contributions to SLA. First, it explores age effects on L3/Ln acquisition, something rarely done in the past. Second, it examines age effects on L3/Ln English grammar by bilingual speakers of those languages (e.g., Urdu and Punjabi) which have not been explored in the past. Third, it uses an editing task, which is a methodological innovation on the pre-existing and excessively used GJTs. Fourth, it compares two age groups in an FL- context whose ages of exposure are similar to those early and late acquirers investigated in SL contexts.

Despite the above stated contributions, this dissertation did not include a baseline assessment of English of the later English MOI group at the end of grade 10, which would be the only direct way to see how much English they might have known before starting English as a MOI. For future studies with a longitudinal design, it would be valuable to examine participants' baseline proficiency to accurately assess the effects of exposure at different ages.

Also, all empirical research to date shows that any effects of the presumed CPH are mediated by the type and amount of exposure to L2. In this study, EL3Ls and LL3Ls had 15,400 and 1,320 hours of exposure to English, respectively. However, data were not collected about the type and quality of exposure. Moreover, using a language as MOI is not the same as naturalistic exposure to an L2.

Another important factor which was not measured in the current dissertation was the role played by participants' aptitude. DeKeyser (2000) and DeKeyser et al. (2010) found a strong, positive, and significant correlation between late learners' scores on GJT and their scores on aptitude test. Similarly Abrahamsson and Hyltenstam (2008) found

that all the L2 learners who performed within native speaker range on GJT obtained an above average aptitude score. Considering the mediating role played by participants' aptitude, future studies should investigate its effects on earlier and later learners' grammaticality scores.

Moreover, that GJTs are only a partial measure of any comprehensive construct of second language proficiency, so results of GJTs must be interpreted in that light. In future, more ecologically valid instruments need to be included in research. For example, a writing task might be used to assess L2 learners' grammatical proficiency. To do this, a coding scheme can be established by analyzing L2 writing for the morphosyntactic features included in influential GJT-based age studies. Doing so might help educators identify specific problematic grammatical features, and, hence, develop materials and direct their teaching to improve such incongruities in L2 learners' writing. Also, findings on such instruments could then be compared with participants' scores on the GJT and editing task.

Additionally, prospective studies might also compare untimed conditions with strictly timed ones. Such a comparison may further enhance our understanding of early and late learners' proficiency differences under such conditions. If late learners perform identically in two conditions, time-stimulus might be excluded as an intervening variable in future research. In contrast, if it shows that the two conditions (i.e., timed and untimed) lead to differential outcomes, SLA researcher might take this into account in future research and include time-stimulus as an important moderating variable among others, such as age of arrival/exposure, age at testing, and length of residence/instruction.

Future studies might also focus on a limited number of grammatical features that might show more sensitivity to age effects for the respective learners under investigation. Also, there should be a sufficient number of items per each grammatical feature to reach an adequate level of reliability on that feature. Additionally, if multiple instruments (GJT, editing task, and writing task) are used, attempts should be made that along with grammatical features (e.g., past tense) all the rule violations for the same feature (e.g., past tense marking omitted, irregular verb regularized, regular ending on irregular stem) should be the same across all the instrument types.

Similar studies in the future might also compare elementary and intermediate groups with a mid-group (i.e., participants who start English medium instruction at the secondary school level). Initially, the current study had a mid-group, which was later excluded because of a relatively smaller number of participants as compared to the other two groups. Findings of this study do not shed any light about the grammatical proficiency of this mid-group, exploration of which might further advance our understanding of language learning in FL contexts.

This study examined Pakistani EL3Ls' and LL3Ls' grammatical ability on a GJT and an editing task; however other areas of language proficiency, such as, pronunciation, pragmatic abilities, reading comprehension, vocabulary depth and breadth, and knowledge and use of collocations were not investigated in the current study. In future along with grammatical ability, research should also be focused on the other abovementioned dimensions of language proficiency.

References

- Abrahamsson, N. (2012). Age of onset and nativelike 12 ultimate attainment of morphosyntactic and phonetic intuition. *Studies in Second Language Acquisition*, *34*, 187-214.
- Abrahamsson, N., & Hyltenstam, K. (2009). Age of onset and nativelikeness in a second language: Listener perception versus linguistic scrutiny. *Language Learning*, *59*, 249-306.
- Asher, J. J., & Garcia, R. (1969). The optimal age to learn a foreign language. *The Modern Language Journal*, 53, 334-341.
- Asher, J. J., & Price. B, S. (1967) The learning strategy of total physical response. *Child Development*, 38, 1219-1227
- Álvarez, E. (2006). Rate and route of acquisition in EFL narrative development at different ages. In C. Muñoz (Ed.), *Age and the rate of foreign language learning* (pp. 127–155). Clevedon: Multilingual Matters.
- Ball, J. F. (1996). *Age and natural order in second language acquisition* (Unpublished Doctoral dissertation). Rochester: University of Rochester.
- Bialystok, E. (1979). Explicit and implicit judgments of L2 grammaticality. *Language Learning*, *29*, 81–103.
- Bialystok, E., & Hakuta, K. (1994). *In other words. The science and psychology of second-language acquisition*. New York, NY: Basic Books.
- Bialystok, E., & Miller, B. (1999). The problem of age in second language acquisition:

 Influences from language, task, and structure. *Bilingualism: Language and Cognition*, 2, 127–145.
- Biostat (2004). What's power analysis? Englewood, NJ: Author.

- Birdsong, D., & Molis, M. (2001). On the evidence for maturational constraints in second-language acquisition. *Journal of Memory and Language*, 44, 235-249.
- Bongaerts, T. (1999). Ultimate attainment in L2 pronunciation: The case of very advanced late L2 learners. In D. Birdsong (Ed.), *Second language acquisition* and the critical period hypothesis (pp. 133-159). Mahwah, NJ: Erlbaum.
- Bott, S. M. (1993). Speech intelligibility and bilingualism: The effects of age of acquisition (Doctoral dissertation, University of Illinois at Urbana-Champaign).

 Retrieved from https://illiad.nau.edu/illiad/.
- Boudreault, P., & Mayberry, R. I. (2006). Grammatical processing in American Sign Language: Age of first-language acquisition effects in relation to syntactic structure. *Language and Cognitive Process*, 21, 608-635.
- Cenoz, J. (2002). Age differences in foreign language learning. *Review of Applied Linguistics*, 13, 125–142.
- Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). New York, NY: Lawrence Erlbaum.
- Curtiss, S. (1977). *Genie: A psycholinguistic study of a modern-day 'wild child.*' New York, NY: Academic Press.
- De Groot, A. M. B. (2011). Language and cognition in bilinguals and multilinguals: An introduction. New York, NY: Psychological Press.
- DeKeyser, R. (2000). The robustness of critical period effects in second language acquisition. *Studies in Second Language Acquisition*, 22, 499-533.

- DeKeyser, R. (2003a). Implicit and explicit learning. In C. Doughty & M. Long (Eds.), *The handbook of second language acquisition* (pp. 313-348). Malden, MA:

 Blackwell.
- DeKeyser, R. (2003b). Confusion about confounding: The critical period and other agerelated aspects of second language learning. Paper presented at ELIA VIII. (Encuentros de Linguistica Inglesa aplicada), "El factor edad en la aquisicion y ensenanza de L2", Seville.
- DeKeyser, R. (2006). A critique of recent arguments against the critical period
 hypothesis. In C. Abello-Contesse, R. Chacón-Beltrán, M. D. López-Jiménez &
 M. M. Torreblanca- López (Eds.), *Age in L2 acquisition and teaching* (pp. 49–58). New York, NY: Peter Lang.
- DeKeyser, R. M. (2013). Age effects in second language learning: Stepping stones toward better understanding. *Language Learning*, *63*, 52-67.
- DeKeyser, R., Alfi-Shabatay, I., & Ravid, D. (2010). Cross-linguistic evidence for the nature of effects in second language acquisition. *Applied Psycholinguistics*, *31*, 413-438.
- Diab, M. N. (2010). Effects of peer- versus self-editing on students' revision of language errors in revised drafts. *System, 38*(1), 85-95.of Language and Education (pp. 119-131). New York, NY: Springer.
- Douglas, D. (2001) Performance consistency in second language acquisition and language testing: A conceptual gap. *Second Language Research* 17, 442456.
- Ellis, N. (2008) Implicit and explicit knowledge about language. In J. Cenoz & N.

- Hornberger (eds.), *Encyclopedia of Language and Education* (pp. 119-131). New York, NY: Springer.
- Ellis, R. (1991). Grammaticality judgments and learner variability. In H. Burmeister & P. Rounds (Eds.), *Variability in second language acquisition: Proceedings of the tenth meeting of the Second Language Acquisition Forum* (pp. 25–60). Eugene, OR: Department of Linguistics, University of Oregon.
- Ellis, R. (2005). Measuring implicit and explicit knowledge of a second language. A psychometric study. *Studies in Second Language Acquisition*, *27*, 141 -172.
- Ellis, R., Loewen, S., Elder, C., Erlam, R., Philp, J., & Reinders, H. (2009). *Implicit and explicit knowledge in second language learning, testing and teaching*. Bristol, UK: Multilingual Matters.
- Evans, S., & Morrison, B. (2011). Meeting the challenges of English-medium higher education: The first-year experience in Hong Kong. *English for Specific Purposes*, 30, 198-208.
- Faul, F., Erdfelder, E., Lang, A, G., & Buchner, A. (2007). G*Power 3: A flexible statistical power analysis program for the social, behavioral, and biomedical sciences. *Behavior Research Methods*, *39*, 175-191.
- Ferris, D., & Roberts, B. (2001). Error feedback in L2 writing classes. How explicit does it need to be? *Journal of Second Language Writing 10*, 161–184.
- Flege, J. E., & Liu, S. (2001). The effect of experience on adults' acquisition of a second language. *Studies in Second Language Acquisition 23*, 527-552.
- Flege, J. E., Yeni-Komshian, G., & Liu, S. (1999). Age constraints on second language learning. *Journal of Memory and Language*, 41, 78–104.

- Francis, N. (2012). *Bilingual competence and bilingual proficiency in child development*.

 Cambridge, MA: MIT Press.
- García Mayo, M. P. (2003). Age, length of exposure, and grammaticality judgments in the acquisition of English as a foreign language. In M.P. García Mayo & M.L. García Lecumberri (Eds.), *Age and the acquisition of English as afForeign language* (pp. 94-114). NY: Multilingual Matters.
- Gass, S. (1983). The development of L2 intuitions. TESOL Quarterly, 17(2), 273-291.
- Granena, G. (2012). Age differences and cognitive aptitudes for implicit and explicit learning in ultimate second language attainment (Doctoral dissertation, University of Maryland, College Park). Retrieved from http://libproxy.nau.edu/docview.
- Granena. G., & Long. M. (2013). Introduction and overview. In G. Garnena & M. Long (Eds.), Sensitive periods, language aptitude, and ultimate L2 attainment (pp. ix-xv). Philadelphia, PA: John Benjamin.
- Gutiérrez, X. (2013). The construct validity of grammaticality judgment tests as measures of implicit and explicit knowledge. *Studies in Second Language Acquisition*, *35*, 423-449.
- Hamid, M, O. (2011). Planning for failure: English and language policy and planning in Bangladesh. In J. Fishman & O. Garcia (Eds.), *Handbook of Language and Ethnic Identity: The success-failure continuum in language and ethnic identity efforts*.
 (pp. 192-203). New York, NY: Oxford University Press.
- Han, Y. (2000) Grammaticality judgment tests: How reliable and valid are they?

 *Applied Language Learning 11, 177204.

- Han, Y., Ellis, R. (1998) Implicit knowledge, explicit knowledge and general language proficiency. *Language Teaching Research* 2, 1-23.
- Harley, B., & Hart, D. (1997). Language aptitude and second language proficiency in Classroom learners of different starting ages. *Studies in Second Language Acquisition*, 19, 379–400.
- Hellman, B. A. (2008). Vocabulary size and depth of word knowledge in adult-onset second language acquisition. *International Journal of Applied Linguistics*, *21*, 162-182.
- Itard, J, M.-G. (1962). *The wild boy of Aveyron (L'Enfant savage)*. New York, NY: Meredith.
- Johnson, J. S. (1992). Critical period effects in second language acquisition: The effect of written versus auditory materials on the assessment of grammatical competence. *Language Learning*, 42, 217-248.
- Johnson, J. S., & Newport, E. (1989). Critical period effects in second language learning:

 The influence of the maturational state on the acquisition of English as a second language. *Cognitive Psychology*, 21, 60-99.
- Johnson, J. S., & Newport, E. (1991). Critical period effects on universal properties of language: The status of subjacency in the acquisition of a second language. *Cognition*, 39, 215-258.
- Kim, R. (1993). A sensitive period for second language acquisition: A reaction time grammaticality judgment task with Korean-English bilinguals. *Issues and Developments in English and Applied Linguistics*, 6, 15-28.

- Krashen, S. D., & Harshman, R. (1972). Lateralization and the critical period. *UCLA Working Papers in Phonetics*, 23, 13-21.
- Krashen, S.D., Long, M.A., & Scarcella, R.C. (1979). Age, rate, and eventual attainment in second language acquisition. *TESOL Quarterly*, *13*, 573–582.
- Larson-Hall, J. (2008). Weighting the benefits of studying a foreign language at a younger starting age in a minimal input situation. *Second Language Research*, 24, 35-63.
- Lenneberg, E. H. (1967). Biological foundations of language. New York, NY: Wiley.
- Linebarger, M. C., Schwartz, M. F., & Saffran, E. M. (1983). Sensitivity to grammatical structure in so-called agrammatic aphasics. *Cognition*, *13*, 361-392.
- Lipsey, M. W., & Wilson, D. B. (2001). *Practical meta-analysis*. Thousand Oaks, CA: Sage.
- Loewen, S. (2009). Grammaticality judgment tests and the measurement of implicit and explicit L2 knowledge. In R. Ellis, S. Loewen, C. Elder, R. Erlam, J. Philp, & H. Reinders (Eds.), *Implicit and explicit knowledge in second language learning, testing and teaching* (pp. 94 112). Bristol, UK: Multilingual Matters.
- Marinova-Todd (2003). Comprehensive analysis of ultimate attainment in adult second language acquisition. *Dissertation Abstract International*, 64 (08), 2814A. (UKI No. 3100151).
- McDonald, J. L. (2000). Grammaticality judgments in a second language: Influences of age of acquisition and native language. *Applied Psycholinguistics*, 21, 395-423.

- McDonald, J. L. (2006). Beyond the critical period: Processing-based explanations for poor grammaticality judgment performance by late second language learners. *Journal of Memory and Language*, 55, 381-401.
- Morgan, G.A., Leech, N.L., Gloeckner, G.W., & Barrett, K.C. (2011). *IBM SPSS for introductory statistics: Use and interpretation* (4th ed.). Mahwah, NJ: Lawrence Erlbaum.
- Moyer, A. (2004). Age, accent and experience in second language acquisition: An integrated approach to critical period inquiry. Clevedon, UK: Multilingual Matters.
- Moyer, A. (2006). Language contact and confidence in L2 listening comprehension: A pilot study of advanced learners of German. *Foreign Language Annals*, *39*, 255–275.
- Moyer, A. (2009). Input as a critical means to an end: Quantity and quality of experience in L2 phonological attainment. In M. Young-Scholten & T. Piske (Eds.), *Input matters in SLA* (pp. 159–174). Bristol, UK: Multilingual Matters.
- Muñoz, C. (2003). Variation in oral skills development and age of onset. In M. P. García Mayo & M. L. García Lecumberri (Eds.), *Age and the acquisition of English as a foreign language: Theoretical issues and fieldwork* (pp. 161-181). Clevedon: Multilingual Matters.
- Muñoz, C. (2006). The effects of age on foreign language learning. In C. Muñoz (Ed.),

 Age and the rate of foreign language learning (pp. 1–40). Clevedon, England:

 Multilingual Matters.

- Muñoz, C. (2008). Age-related differences in foreign language learning. Revisiting the empirical evidence. *International Review of Applied Linguistics in Language Teaching*, 46, 197–220.
- Muñoz, C. (2011). Input and long-term effects of starting age in foreign language learning. *International Review of Applied Linguistics in Language Teaching*, 71, 197-220.
- Muñoz, C., & Singleton, D. (2007). Foreign accent in advanced learners: Two successful profiles. *The EUROL2A Yearbook*, *7*, 171–190.
- Navés, T., Torras, M. R., & Celaya, M. (2003). Long-term effects of an earlier start: An Analysis of EFL written production. *EUROSLA Yearbook*, *3*, 103-129.
- Oswald, F. L., & Plonsky, L. (2010). Meta-analysis in second language research: Choices and challenges. *Annual Review of Applied Linguistics*, 30, 85-110.
- Patkowski, M. (1980). The sensitive period for the acquisition of syntax in a second language. *Language Learning*, *30*, 449–472.
- Purpura, J. (2004) Assessing grammar. Cambridge: Cambridge University Press.
- Qureshi, M. A. (2015). Age and ultimate attainment in a second language: A meta-analysis. Manuscript submitted for publication.
- Schachter, J., & Yip, V. (1990). Grammaticality judgments: why does anyone object to subject extraction? *Studies in Second Language Acquisition 12*: 379-392.
- Seol, H. (2005). The critical period in the acquisition of L2 syntax: A partial replication of Johnson and Newport, *Educational Psychology*, 24, 77-97.

- Shim, R. J. (1993). Sensitive period for second language acquisition: A reaction time study of syntactic processing of English by Korean-English bilinguals. *Ideas and Developments in English and Applied Linguistics*. *6*, 43-64.
- Singleton, B., & Muñoz, C. (2011). Around and beyond the critical period. In E. Hinkel (Ed.), *Handbook of research in second language teaching and research*. (vol. II), (pp. 407-425). New York, NY: Routledge.
- Slavoff, G. R., & Johnson, J. S. (1995). The effects of age on the rate of learning a second language. *Studies in Second Language Acquisition*, 17, 1–16.
- Snow, C., & Hoefnagel-Hohle, M. (1978). The critical period for language acquisition: Evidence from second language learning. *Child Development*, 49, 1114–1128.
- Spada, N. (2015). SLA research and L2 pedagogy: Misapplications and questions of relevance. *Language Teaching*, 48 (1), 69—81.
- Storch, N. (2002). Patterns of interaction in ESL pair work. *Language Learning* 5(1), 119–158.
- Storch, N. (2007). Investigating the merits of pair work on a text-editing task in ESL classes. *Language Teaching Research*, 11(2), 143-159.
- Torras, M., & Celaya, M. (2001). Age-related differences in the development of written production. An empirical study of EFL school learners. *International Journal of English Studies*, 1, 103-126.
- Vogt, W. P. (1999). Dictionary of statistics & methodology: A nontechnical guide for the social sciences. Thousand Oak, CA: Sage

Appendix A

Script for Consent Form

Note: Items in brackets [] are not to be read aloud.

[After participants have been seated, the researcher will say the following]

As'salam 'alykum, I am Asif, a Ph.D. student in Applied Linguistics. I am really grateful for your time and cooperation in my study. For my Ph.D. dissertation, I am investigating age effects on early and late third language acquisition by comparing people who were exposed to English at elementary, secondary, and intermediate levels. The research question that I want to answer is this: Does exposure to English at different levels of education (i.e., elementary, secondary, and intermediate) affect later language proficiency?

To answer this question, I need your help. I have two main tasks for you: a grammaticality judgment test and an editing task. In addition to these two main tasks, I am asking you to complete an informed consent form and a language background questionnaire. I will explain each part to you. We will start with an explanation of the informed consent form.

I will pass out the informed consent forms now. You don't need to read the forms yet, just put your copy face down [demonstrate].

[After all the participants have received an informed consent form]

Let's look at the form together. I will read it aloud, while you read along.

[After I finish reading the consent form, I will say the following]

Do you have any questions about the content of the form?

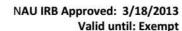
[Any questions/confusions will be addressed.]

Turn to page 2. At the bottom of the page, please sign your name, write your name, and write today's date.

Afterwards, consent forms will be collected from all the students. This entire process might take approximately 10 minutes. The consent form is in Appendix B.

Appendix B

Informed Consent Form



Page 1 of 2





Human Subject Informed Consent English Department 700 S. Humphreys CDR, Flagstaff, AZ, 86011. (928) 523-4911

Project Title: Age and knowledge of English: Intuition, production, and error correction

Dear Participant:

You are being asked to participate in a research project conducted through the **Department of English** at Northern Arizona University by **Muhammad Asif Qureshi**. The researcher is required to receive your informed consent before you participate in this project.

Muhammad Asif Qureshi will explain to you in detail: (1) the purpose of the project; (2) what you will be asked to do and how long your participation will last; (3) how your personal information, if collected, will be kept confidential; (4) if you will receive any compensation; (5) the benefits; and (6) potential risks of participation.

Your participation in research is voluntary. If you choose not to participate, there are no penalties or loss of benefits or services that you are otherwise entitled. If you decide to participate and then withdraw or skip a question, there are also no penalties or loss of benefits or services. Whether or not you choose to participate in this project will have no effect on your relationship with NAU now or in the future.

A basic explanation of the project appears below. Please read this explanation and discuss it with **Mr. Qureshi.** Feel free to ask questions to help you understand the project. After any questions you may have are answered and you decide to participate in the research, please sign on the last page of this form in the presence of the person who explained the project to you. A copy of this form will be given to you to keep.

1. PROJECT PURPOSE:

The study investigates the possible influence of age on knowledge of English structures.

2. EXPLANATION OF PROCEDURES:

- In this research study, you will be asked to:
- 1) Sign a consent form,
- 2) Fill out a Background Questionnaire (BQ), take a Grammar Judgment Test (GJT), write an argumentative essay, and edit a written task. The overall duration of all activities will be an hour and twenty minutes.

3. CONFIDENTIALITY:

Your identity will be kept strictly confidential. Your name will be removed from the background questionnaire and will be replaced with an anonymous code. The records of this study will be kept private. Any written results will discuss group findings and will not include information that will identify you.

4. COMPENSATION:

Participation in the study is voluntary and no compensation will be paid to the participants.

NAU IRB Approved: 3/18/2013 Valid until: Exempt

Page 2 of 2



5. BENEFITS:

There are no immediate benefits to individual participants. However, their universities may be provided with the results of the study which might help them understand their students' proficiency in English.

6. RISKS:

There are no known risks associated with this project other than those encountered in daily life.

7. CONSENT:

I have read the above information about **AGE AND KNOWLEDGE OF ENGLISH: INTUITION, PRODUCTION AND ERROR CORRECTION** and have been given an opportunity to ask questions. I agree to participate in this project, and I have been given a copy of this consent document.

Signature of Participant	Date
Printed Name of Participant	
Signature of Research Representative	Date
Printed Name of Research Representative	

The dated approval stamp in the header of this consent form indicates that this project has been reviewed and approved by the Northern Arizona University Institutional Review Board (IRB) for the Protection of Human Subjects in Research. Contact the Human Research Protections Office at 928-523-4236 if you have any questions about: (1) the conduct of the project, or (2) your rights as a research participant, or (3) a research-related injury. Any other questions about the research project should be directed to:

Muhammad Asif Qureshi

maq7@nau.edu 646-705-6838

Dr. Mary McGroarty

Department of English mary.mcgroarty@nau.edu 928-523-6273

Dr. Luke Plonsky

Department of English luke.plonsky@nau.edu 928-523-6265

Appendix C

Script for the Background Questionnaire

Note: Items in brackets [] are not to be read aloud.

Thanks for signing the informed consent form and agreeing to participate in the study. Now, let's continue with an explanation of the language background questionnaire.

The language background questionnaire has 17 questions that inquire about your educational background, your age when you were exposed to English as a third language, your current age, your self-rated English proficiency, and the number of languages you know.

You have 10 minutes to complete the background questionnaire. After 7 minutes, I will remind you that you have 3 minutes left [show 3 Minutes Remaining sign]. After 10 minutes, I will ask you to stop [show Stop sign].

Background Questionnaire is in Appendices D.

Appendix D

Language Background Questionnaire

Thank you very much for participating in this study. This language background questionnaire contains questions concerning your experiences with Urdu, Punjabi, English, and other languages. I deeply appreciate your cooperation in filling out the questionnaire as accurately as possible. You have **10 minutes** to complete the background questionnaire.

Fill in the blanks or check () the box, where appropriate.

1.	What is your year of birth? (e.	g. 1973). , fie	ld of study						
			?						
2. 1	am Male Female								
3. (Currently, I am an undergradu	ate student. Yes 🗌 No 🗌							
4.]	4. If your answer to #3 is yes, what year are you in? 1 st 2 nd 3 rd 4 th								
5. (Currently, I am an M.A or M.	Sc student. Yes No							
6. \	What language/languages do y	you speak at home?							
	1								
7.]	How old were you when you s	started speaking Urdu?							
8. 1	How old were you when you s	started speaking Punjabi?							
9. '	What do you consider as your	second language?							
10.	How old were you when you	started speaking your seco	ond language?						
11.	What other languages do you	ı know?							
12.	Readings and class assignme	ents, in classes other than U	rdu and Arabic, were in						
F	English at my								
	primary school level	Yes 🗌	No 🗔						
-	secondary school level	Yes 🗌	No 🗌						
	intermediate level	Yes 🗌	No 🗌						
	Undergraduate level	Yes 🗌	No 🗌						
	Master's degree level	Yes 🗌	No 🗌						

13. If the medium of instruction in your primary school was not English, did you										
attend any	cour	rses that had En	ıglish as a subj	ect? Yes \square N	0					
14. If y	es, for h	ow many hours	w many hours a day , and for how many days							
every we	ek	?								
15. If the r	15. If the medium of instruction in your secondary school was not English, did you attend									
any o	courses th	nat had English	as a subject?	Yes No						
16. If yes,	16. If yes, for how many hours a day, and for how many days every									
week		?								
17. Have y	ou ever	visited an Engl	ish speaking co	ountry and stay	ed there for m	ore than 3				
month	s? Yes	No 🗌								
18. Please	provide	the number of	years	and	months					
you sta	ayed in an	n English speak	king country.							
19. Did yo	u ever at	tend an English	n language cen	ter in Pakistan'	? Yes \[\] No [
20. What v	was your	age when you	enrolled in the	English langua	age center?					
21. For ho	w many	years	, months	, day	s per week	,				
and ho		_	day did you at							
	_	w much time do	o you spend ev	ery day watchi	ing the followi	ng in				
Englis										
	channels		glish films		shows					
23. I consi	der my p	resent English	language profi	ciency to be						
		Poor	Below	Average	Above	Very good				
			average		average					
List	ening									
Rea	ading									
Wr	iting									
Spe	eaking									

Thank you so much for completing the Background Questionnaire.

Appendix E

Table of Specifications for Grammaticality Judgment Test

Participants: Early and late EFL learners in Pakistan
Educational context: Undergraduate and graduate students
Use of test: Evaluation of morphosyntactic proficiency

Response type(s): Judging grammaticality as correct or incorrect

Scoring: Dichotomous scoring: correct response = 1, incorrect response = 0

Time: 25 minutes

Features	Pst.T	Plul	T.Pr	Prog	Detr	Pron	Pti.m	Subc	Y/No	Wh	W.O	T. Itms	# Pnts
# of rule types ²	3	3	2	2	2	2	2	1	4	2	5		
# of items per rule-	4	4	4	4	4	4	4	20	4	4	4		
type Total # of items	12	12	8	8	8	8	8	20	16	8	20	128	128
% of items per rule type.	9.30	9.30	6.25	6.25	6.25	6.25	6.25	15.6	12.50	6.25	15.6		

Note: $Pst.T = Past \ tense, \ Plul = Plurals, \ T.Pr = Third \ Person, \ Prog = Progressive, \ Detr = Determiners, \ Pron = Pronominalization, \ Pti. \ M = Particle \ Movement, \ Subc = Subcategorization, \ Y.NO = Yes/No \ Questions, \ Wh = WH-questions, \ WO = Word \ Order, \ T. \ Itms = Total \ number \ of \ items, \ \# Pnts = Total \ points$

² Rule-types refers to the type of error, for example past tense has errors related to three rule types (i.e., past tense marking omitted, irregular verb regularized, and regular ending on irregular stem).

Appendix F

Script for Grammaticality Judgment Test

Note: Items in brackets [] are not to be read aloud.

Now we will turn to the last task of the day, the grammaticality judgment test. The test contains 128 sentences. Half of the sentences, which means 64 sentences, have one grammatical error, for example, the wrong use of the past tense. The remaining 64 sentences are grammatically correct. After each sentence, two boxes are provided. One box is to indicate a correct sentence (that is, a grammatical sentence) and the other box is to indicate an incorrect sentence (that is, an ungrammatical sentence). Your job, as you complete the test, is to determine whether a sentence is grammatical or not. Place a check mark (\checkmark) in the appropriate box to signal if each sentence is grammatical or not. Let's look at four examples now. [If a projector is available, participants will be shown examples on the screen, Otherwise, I will write examples on the board. Examples are in Appendix P].

[After participants are shown the examples, GJT test-copies will be distributed.

[Then they will be told the following,]

You have 25 minutes to complete this test. After 20 minutes, I will remind you about the time by raising a sign showing that you have only 5 minutes left [I will display the sign here]. After 25 minutes, I will stop you, again by raising a stop sign [I will display the stop sign here]. Then, I will collect the GJT task copies from you.

Now the time is [time will be stated and written on the board]. You can begin your test now. [After I have said this, I will write the exact time on the board and tell participants that they can begin the test. After 20 minutes, participants will be informed about the time.

After 25 minutes they will be requested to stop.]

Examples of the GJT, the GJT task for the participants, and the GJT with answers are in Appendices G, H, and I, respectively.

Appendix G

Examples for the Grammaticality Judgment Task

Practice Items	Correct	Incorrect
1. A snake bit she on the leg.		✓
2. Susan is making some cookies for us.	V	
3. The baby bird has fall from the oak tree.		V
4. The little boy was counting all his pennies last	V	
night.		

Appendix H

Grammaticality Judgment Task

Please read each sentence. Decide if the sentence is correct (grammatical) or incorrect (ungrammatical). Place a check mark () in the appropriate box after each sentence.

You have 25 minutes to complete this test.

Note: Some sentences will look similar to others. Read each sentence carefully and then respond.

After you finish the test, please remain seated. All the tests will be collected at the end the session, which means after 25 minutes.

No	Actual Items	Correct	Incorrect
	Last night the old lady die in her sleep.		
4	Yesterday the teacher sented Allison to the principal.		
(Janie sleeped with her teddy bear last night.		
4	A policeman gived Alan a ticket for speeding yesterday.		
:	I want you to go to the store now.		
(John said me that his wife was ill.		
,	The student was learning in his room until late last night.		
	I want you will go to the store now.		
9	I hope you will go to the store now.		
	The man allows his son to watch TV.		
	The little boys laughed at the clown.		
	John's dog always waits for him at the corner.		
	Mr. Murphy hidded his money under his mattress.		
	Many house were destroyed by the flood last week.		
	Two mouses ran into the house this morning.		
	Many houses were destroyed by the flood last week.		
	The farmer bought two pig at the market.		
	Mary will go to Europe next year.		
	John's dog always waits for him at the corner.		
1	I need to get some information about the train schedule.		
1	The farmer bought two pigs at the market.		
1	A shoe salesman sees many feet throughout the day.		
,	Two mice ran into the house this morning.		
2	Our neighbor bought new furniture last week.		
1	Every Friday our neighbor washes her car.		
1	I need to get some informations about the train schedule.		
1	Our neighbor bought new furnitures last week.		
,	John's dog always wait for him at the corner.		
	What is Martha bringing to the party?		
	Mary will goes to Europe next year.		
	The Johnsons may are moving to Chicago this fall.		
	The children playing in the garden till dark these days.		

Bob is trying to fix Jim's car with his new tools.		-
de la constantina del constantina de la constantina de la constantina del constantina de la constantin	Correct	Incorrect
The little boy is speak to a policeman.		
The Johnsons may be moving to Chicago this fall.		
Janet is wearing the dress I gave her.		
Bob trying to fix Jim's car with his new tools.		
Tom is reading a book in the bathtub.		
Janet is wear the dress I gave her.		
The little boy is speaking to a policeman.		
The children are playing in the garden till dark these days.		
Beauty is something that lasts forever.		
Tom is reading book in the bathtub.		
Mrs. Johnson went to library yesterday.		
The man looked the new cars over yesterday.		
The new neighbors carried a long conversation on.		
The beauty is something that lasts forever.		
The girl cut herself on a piece of glass.		
The new neighbors carried on a long conversation.		
Mrs. Johnson went to the library yesterday.		
After a life like that he will go straight to the hell.		
Mike wrote the letter but didn't send.		
Peter made out the check but didn't sign.		
The girl cut himself on a piece of glass.		
John knew but she did not tell.		
Will Harry be blamed for the accident?		
Has the King been served his dinner?		
The man climbed the ladder up carefully.		
George says much too softly.		
The girls enjoy to watch TV.		
I hope you to go to the store now.		
She took her coat off quickly.		
The man looked the new cars yesterday over.		
After a life like that he will go straight to hell.		
Peter made out the check but didn't sign it.		
Sandy filled a jar with cookies last night.		
Last night the old lady died in her sleep.		
A policeman gave Alan a ticket for speeding yesterday.		
Mr. Murphy hid his money under his mattress.		
Yesterday the teacher sent Allison to the principal.		
The man allows his son watch TV.		
A shoe salesman sees many foots throughout the day.		
Janie slept with her teddy bear last night.		
John told me that his wife was ill.		
Will wear Harry his new shirt to the party?		
Did Bill dance at the party last night?		

Is waiting Sally in the car?		_
	Correct	Incorrect
The little boys laughed the clown.		
The student was studying in his room until late last night.		
She took her coat quickly off.		
The man lets his son watch TV.		
John knew but he did not tell.		
Does Martha use her microwave oven?		
The girls want watching TV.		
Will be Harry blamed for the accident?		
The man climbed up the ladder carefully.		
The man lets his son to watch TV.		
Has been the King served his dinner?		
The girls want to watch TV.		
Mike wrote the letter but didn't send it.		
Will Harry wear his new shirt to the party?		
Linda baked John a cake.		
Every Friday our neighbor wash her car.		
Did Bobbie stay at home last night?		
Danced Bill at the party last night?		
Swam Janet in the race yesterday?		
What Martha is bringing to the party?		
When will Sam fix his car?		
The girls enjoy watching TV.		
Is Sally waiting in the car?		
Who do you meet at the park every day?		
The woman the policeman asked a question.		
When do they leave for Mexico?		
The woman asked the policeman a question.		
Did Bobbie stayed at home last night?		
Does Martha uses her microwave oven?		
The boy caught the ball.		
The girl likes the movie.		
When Sam will fix his car?		
When they leave for Mexico?		
The ball the boy caught.		
The girl the movie likes.		
Who you meet at the park every day?		
Linda a cake baked John.		
The student eats his meals quickly.		
All our friends live in the suburbs.		
Kevin usually rides his bicycle to work.		
The man drinks.		
The children play with the dog.		
The student eats quickly his meals.		

Paints the woman.		
	Correct	Incorrect
Did Janet swim in the race yesterday?		
Drinks the man.		
George says his prayers much too softly.		
The woman paints.		
The children with the dog play.		
All our friends in the suburbs live.		
Kevin rides usually his bicycle to work.		

Appendix I

Grammaticality Judgment Task with Answers

No	Actual Items	Correct	Incorrect
	Last night the old lady die in her sleep.		V
1	Yesterday the teacher sented Allison to the principal.		V
	Janie sleeped with her teddy bear last night.		V
4	A policeman gived Alan a ticket for speeding yesterday.		✓
,	I want you to go to the store now.	/	
	John said me that his wife was ill.		V
,	The student was learning in his room until late last night.		V
	I want you will go to the store now.		✓
9	I hope you will go to the store now.	/	
	The man allows his son to watch TV.	/	
	The little boys laughed at the clown.	/	
	John's dog always waits for him at the corner.	/	
	Mr. Murphy hidded his money under his mattress.		'
	Many house were destroyed by the flood last week.		/
	Two mouses ran into the house this morning.		V
	Many houses were destroyed by the flood last week.	/	
	The farmer bought two pig at the market.		/
	Mary will go to Europe next year.	/	
	John's dog always waits for him at the corner.	/	
	I need to get some information about the train schedule.	/	
	The farmer bought two pigs at the market.	/	
	A shoe salesman sees many feet throughout the day.	/	
	Two mice ran into the house this morning.	/	
2	Our neighbor bought new furniture last week.	V	
	Every Friday, our neighbor washes her car.	/	
	I need to get some informations about the train schedule.		'
	Our neighbor bought new furnitures last week.		/
	John's dog always wait for him at the corner.		'
	What is Martha bringing to the party?	/	
-	Mary will goes to Europe next year.		V
	The Johnsons may are moving to Chicago this fall.		V
-	The children playing in the garden till dark these days.		/
(Bob is trying to fix Jim's car with the new tools.	/	-
(The little boy is speak to a policeman.	-	V
	The Johnsons may be moving to Chicago this fall.	/	-

	Correct	Incorrect
Janet is wearing the dress I gave her.	/	
Bob trying to fix Jim's car with the new tools.		✓
Tom is reading a book in the bathtub.	/	
Janet is wear the dress I gave her.		/
The little boy is speaking to a policeman.	/	
The children are playing in the garden until dark these days.	/	
Beauty is something that lasts forever.	/	
Tom is reading book in the bathtub.		V
Mrs. Johnson went to library yesterday.		V
The man looked the new cars over yesterday.		V
The new neighbors carried a long conversation on.		V
The beauty is something that lasts forever.		'
The girl cut herself on a piece of glass.	/	
The new neighbors carried on a long conversation.	/	
Mrs. Johnson went to the library yesterday.	/	
After a life like that, he will go straight to the hell.		/
Mike wrote the letter but didn't send.		/
Peter made out the check but didn't sign.		'
The girl cut himself on a piece of glass.		/
John knew but she did not tell.		/
Will Harry be blamed for the accident?	/	
Has the King been served his dinner?	/	
The man climbed the ladder up carefully.	-	/
George says much too softly.		/
The girls enjoy to watch TV.		/
I hope you to go to the store now.		/
She took her coat off quickly.	/	
The man looked the new cars yesterday over.	-	/
After a life like that, he will go straight to hell.	/	
Peter made out the check but didn't sign it.	V	
Sandy filled a jar with cookies last night.	/	
Last night the old lady died in her sleep.	/	
A policeman gave Alan a ticket for speeding yesterday.	/	
Mr. Murphy hid his money under his mattress.	V	
Yesterday the teacher sent Allison to the principal.	· ·	
The man allows his son watch TV.		✓
A shoe salesman sees many foots throughout the day.		~
Janie slept with her teddy bear last night.	/	•
John told me that his wife was ill.	1	

	Correct	Incorrect
Will wear Harry his new shirt to the party?		/
Did Bill dance at the party last night?	/	
Is waiting Sally in the car?		/
The little boys laughed the clown.		/
The student was studying in his room until late last night.	✓	
She took her coat quickly off.		✓
The man lets his son watch TV.	/	
John knew but he did not tell.	✓	
Does Martha use her microwave oven?	/	
The girls want watching TV.		V
Will be Harry blamed for the accident?		V
The man climbed up the ladder carefully.	V	
The man lets his son to watch TV.		~
Has been the King served his dinner?		/
The girls want to watch TV.	/	
Mike wrote the letter but didn't send it.	/	
Will Harry wear his new shirt to the party?	/	
Linda baked John a cake.	/	
Every Friday our neighbor wash her car.		V
Did Bobbie stay at home last night?	/	
Danced Bill at the party last night?		'
Swam Janet in the race yesterday?		/
What Martha is bringing to the party?		/
When will Sam fix his car?	/	
The girls enjoy watching TV.	V	
Is Sally waiting in the car?	/	
Who do you meet at the park every day?	<i>\</i>	
The woman the policeman asked a question.		/
When do they leave for Mexico?	/	
The woman asked the policeman a question.	· /	
Did Bobbie stayed at home last night?		1
Does Martha uses her microwave oven?		1
The boy caught the ball.		
The girl likes the movie.	1	
When Sam will fix his car?		~
When they leave for Mexico?		1
The ball the boy caught.		1
The girl the movie likes.		✓
Who you meet at the park every day?		

	Correct	Incorrect
Linda a cake baked John.		V
The student eats his meals quickly.	V	
All our friends live in the suburbs.	V	
Kevin usually rides his bicycle to work.	V	
The man drinks.	V	
The children play with the dog.	V	
The student eats quickly his meals.		✓
Paints the woman.		✓
Did Janet swim in the race yesterday?	V	
Drinks the man.		V
George says his prayers much too softly.	V	
The woman paints.	V	
The children with the dog play.		'
All our friends in the suburbs live.		/
Kevin rides usually his bicycle to work.		~

Appendix J

Table of Specifications for Editing Task

Participants: Early and late EFL learners in Pakistan.
Educational context: Undergraduate and graduate students.
Use of test: Evaluation of morphosyntactic proficiency.

Response type(s): Identifying morphosyntactic errors and providing corrections.

Scoring: Dichotomous scoring: correct response = 1, incorrect response or no change = 0

Time: 20 minutes

Features	Pst.T	Plul	T.Pr	Prog	Detr	Pron	Pti.m	Subc	Aux	W.O	Adv	T. Itms	# Pnts
# of rule-types ³	1	1	1	1	1	1	1	2	1	1	1		
# of items per rule-type	2	2	2	2	2	2	2	4	2	2	2		
Total # of items	2	2	2	2	2	2	2	4	2	2	2	24	24
% of items per rule type	8.3	8.3	8.3	8.3	8.3	8.3	8.3	16.6	8.3	8.3	8.3	8.3	

Note: $Pst.T = Past \ tense, \ Plul = Plurals, \ T.Pr = Third \ Person, \ Prog = Progressive, \ Detr = Determiners, \ Pron = Pronominalization, \ Pti. \ M = Particle \ Movement, \ Subc = Subcategorization, \ Aux = Auxiliaries, \ W.O = Word \ Order. \ Adv = Adverb, \ T. \ Itms = Total \ number \ of \ items, \ \# Pnts = Total \ points$

³ Rule-types refers to the type of error, for example, sub-categorization has errors related to two rule types (i.e., gerunds and infinitives).

Appendix K

Script for Editing Task

Note: Items inside brackets [] are not to be read aloud.

Well, it seems like everything is going fine. Thanks for finishing the background questionnaire! Now we will start the first major task, the editing task. I will be distributing a short passage of 222 words. The passage has 24 grammatical errors in it. Like an editor, you need to identify each error, cross it out, and then correct it in the blank space below the error. There are different kinds of errors. In certain cases, for example, you might have to rearrange the word order in a sentence. Now I will show you some examples to help you understand the editing task. [If a projector is available, participants will be shown examples on the screen, otherwise, I will write examples on the board].

[After participants have seen the examples, copies of the editing task will be distributed. Then I will say the following,]

Do you have any questions about the editing task?

[Any questions/confusions will be addressed. Then they will be told the following,]

You have 20 minutes to complete this task. After 15 minutes, I will remind you about the time by raising a sign showing that you have only 5 minutes left [I will display the sign here]. After 20 minutes, I will stop you, again by raising a stop sign [I will display the stop sign here]. Then, I will collect the editing task copies from you.

[After I have said this, I will write the exact time on the board and tell participants that they can begin the test. After 15 minutes, participants will be informed about the time.

After 20 minutes they will be requested to stop.]

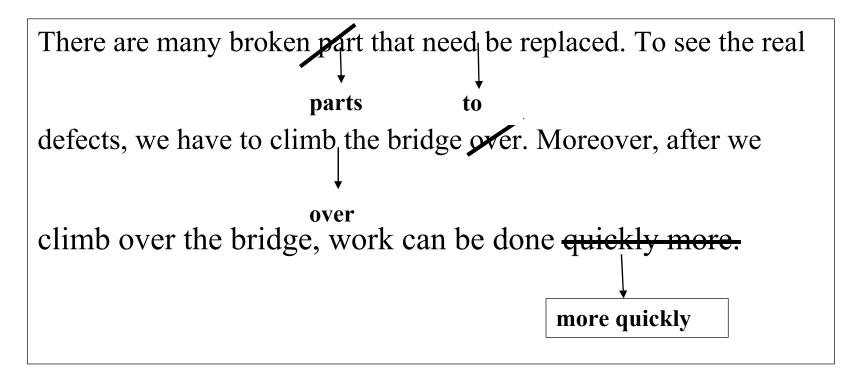
Examples of the editing task, the editing task for the participants, the editing task with answers in bold, and table of potential answers for the editing task are in Appendices L, M, N, and O, respectively

Appendix L

Example of Editing Task

Directions: Correct/edit the following text for grammatical accuracy. While editing/correcting the errors, you might need to do one of the following: (a) cross an error and replace it with the correct form, (b) rearrange word order in few sentences (which could also involve crossing out something), and (c) insert a missing word. During editing/correcting, please provide the correct form in the empty space provided <u>below</u> each sentence.

Example:



Appendix M

The Editing Task

Directions: Correct/edit the following text for grammatical accuracy. While editing/correcting the errors, you might need to do one of the following: (a) cross an error and replace it with the correct form, (b) rearrange word order in few sentences (which could also involve crossing out something), and (c) insert a missing word. During editing/correcting, please provide the correct form in the empty space provided <u>below</u> each sentence.

could also involve crossing out something), and (c) insert a missing word. During editing/correcting, please provide the correct form in the empty space provided <u>below</u> each sentence.
Before you begin, please note down the starting time (e.g., 9:15). When you finish the editing task, please mark finishing time the space provided at the end of the editing task
There are serious problems in many part of Africa. The biggest problem face
Africans today is the continue threat of wars. In the past, the fighting was local and small
scale. Now it has become far more damaging. It is because the situation has changed
dramatic.
European powers begin to move into Africa in sixteenth century. They have taken
African people to sell as slaves in North and South America. They also tooks any
valuable resources they could find. They were not understand much about African tribal

tradition, and so the borders of these countries not match traditional borders of tribal lands.

Many of the problems facing today Africa have been worsened by fighting control over of the government. Countries that are at war have less time or resources to deal with poverty, hunger, or disease. They are unable take any measures for a better future, and so many countries are become poorer and their problems are grow. For example, HIV, virus that cause AIDS, rapidly has spread in Africa. Other diseases have also spread quick.

In spite of these problems, however, many Africans are hopeful about future.

Young and talented Africans are looking to the rest of the world out. They are

experimenting new ways to use the internet and other technology to try to solving some

of her problems.

Time Finished:	 	
Total time:		

Appendix N

Editing Task with Answers in Bold

1	There are serious problems in many parts of Africa. The biggest problem facing
2	Africans today is the continuing/continued threat of wars. In the past, the fighting was local and
3	small scale. Now it has become far more damaging. It is because the situation has changed
4	dramatically.
5	European powers began to move into Africa in sixteenth century. They took African
6	people to sell as slaves in North and South America. They also took any valuable resources they
7	could find. They did not understand much about African tribal traditions , that's why the borders
8	of these countries do not match the traditional borders of tribal lands.
9	Many of the problems facing Africa today have been worsened by fighting over control
10	of the government. Countries that are at war have less time or resources to deal with poverty,
11	hunger, or disease. They are unable to take any measures for a better future, as a result many

12	countries are becoming/have become poorer and their problems are growing/have grown . For
13	example, HIV, the virus that causes AIDS, has rapidly spread in Africa. Other diseases have
14	also spread quickly.
15	In spite of these problems, however, many Africans are hopeful about their future.
16	Young and talented Africans are looking out to the rest of the world. They are experimenting

new ways to use the internet and other new technology to try to **solve** some of **their** problems.

Appendix O

Potential Answers for Editing Task

Note: Answers other than those provided in Appendices J and L can also be considered as correct if two raters agree on them.

Features Line		Error and Correction Line		Error and	
	#		#	Correction	
Past tense	5	Begin (began)	5	have taken (took)	
Plural	1	part (parts)	7	tradition (traditions)	
Third Person	6	tooks (took)	13	cause (causes)	
Present Progressive	12	become (ing/have become)	12	grow (growing/have grown)	
Determiners	8	traditional (the traditional)	13	virus (the virus)	
Pronominalization	15	future (their future)	17	her (their)	
Particle Movement	9	fighting control over (fighting over control)	19	looking out (looking out/looking)	
Subcategorization					
Gerunds	1	face (facing)	2	continue (ing/ed)	
Infinitive	11	take (to take)	17	to solving (to solve)	
Aux. wrong/omitted	7	were not (did not)	8	not (do not)	
Word Order	9	facing today Africa (facing Africa today)	13	rapidly has (has rapidly)	
Adverb	4	dramatic (dramatically)	14	quick (quickly)	

Total number of items, # Pnts = Total points

Appendix P

Coding Guide for the Editing Task

1	Past tense, lines 1 &	& 2, Errors: (a) begin, (b) have taken.
	Sample sentence 1 Correction 1 Sample sentence 2 Correction 2	European powers begin to move into Africa in sixteenth century. European powers began to move into Africa in sixteenth century. They have taken African people to sell as slaves in North and South America. They took African people to sell as slaves in North and South America century.
2	Plurals, lines 1 &	7, Errors: (a) part, (b) tradition
	Sample sentence 1 Correction 1 Sample sentence 2 Correction 2	There are serious problems in many parts of Africa There are serious problems in many parts of Africa They did not understand much about African tribal tradition, that's why the borders of these countries do not match the traditional borders of tribal lands. America. They did not understand much about African tribal traditions, that's why the borders of these countries do not match the traditional borders of tribal lands. America.
3	Third person, lines	s 6 & 13, Errors: (a) tooks, (b) cause
	Sample sentence 1 Correction 1 Sample sentence 2 Correction 2	They also tooks any valuable resources they could find. They also took any valuable resources they could find. HIV, the virus that cause AIDS, has rapidly spread in Africa. HIV, the virus that causes AIDS, has rapidly spread in Africa.
4	4. Present progress	sive, lines 12 & 12, Errors: (a) are become, (b) are grow.
	Sample sentence (error 1 & 2) Correction a (error 1 & 2) Correction 1b	They are unable to take any measures for a better future, as a result many countries are become poorer and their problems are grow. They are unable to take any measures for a better future, as a result many countries are becoming poorer and their problems are growing . They are unable to take any measures for a better future, as a result many countries are poor/er and their problems are growing.
	Correction 1c & 2b Incorrect	They are unable to take any measures for a better future, as a result many countries have become poorer and their problems have grown . They are unable to take any measures for a better future, as a result many countries are poor and their problems are grown .

5	Determiners, lines	8 & 13, Errors: (a) traditional, (b) virus
	Sample sentence 1	They did not understand much about African tribal traditions, that's why the
	Correction 1	borders of these countries do not match traditional borders of tribal lands. They did not understand much about African tribal traditions, that's why the borders of these countries do not match the traditional borders of tribal lands.
	Sample sentence 2 Correction 2	For example, HIV, virus that causes AIDS, has rapidly spread in Africa. For example, HIV, the virus that causes AIDS, has rapidly spread in Africa.
6	Pronominalization	, lines 15 & 17, Errors: (a) about future, (b) her problems
	Sample sentence 1	In spite of these problems, however, many Africans are hopeful about future.
	Correction 1a	In spite of these problems, however, many Africans are hopeful about their future.
	Correction 1b	In spite of these problems, however, many Africans are hopeful about the future.
	Sample sentence 2	They are experimenting new ways to use the internet and other new technology to try to solve some of her problems.
	Correction 2a	They are experimenting new ways to use the internet and other new technology to try to solve some of their problems.
	Correction 2b	They are experimenting new ways to use the internet and other new technology to try to solve some of her problems.
7	Particle movement	, lines 9 & 16, Errors: (a) fighting control over, (b) lookingout
	Sample sentence 1	Many of the problems facing Africa today have been worsened by fighting control over of the government
	Correction 1a	Many of the problems facing Africa today have been worsened by fighting over control of the government
	Correction 1b	Many of the problems facing Africa today have been worsened by fighting for control of the government
	Sample sentence 2 Correction 2a	Young and talented Africans are looking to the rest of the world out. Young and talented Africans are looking out to the rest of the world.
	Correction 2b	Young and talented Africans are looking to the rest of the world out.
	Correction 2c	Young and talented Africans are looking at the rest of the world.
8	Gerunds, lines 1 &	2, Errors: (a) problems face Africans (b) is the continue threat
	Sample sentence 1	The biggest problem face Africans today is the continuing threat of wars.
	Correction 1a	The biggest problem facing Africans today is the continuing threat of wars.
	Correction 1b	The biggest problem faced by Africans today is the continuing threat of wars.
	Correction 1c	The biggest problem Africans face today is the continuing threat of wars.
	Correction 1d	The biggest problem which Africans face today is the continuing threat of war.
	Correction 1e	The biggest problem, which Africans are facing today, is the continuing

	threat of war.
Incorrect	The biggest problem faced Africans today is the continuing threat of wars.
Sample sentence 2	The biggest problem facing Africans today is the continue threat of wars.
Correction 2a	The biggest problem facing Africans today is the continuing threat of wars
Correction 2b	The biggest problem facing Africans today is the continued threat of wars.
Correction 2c	The biggest problem facing Africans today is the continuous threat of wars

9 Infinitives, lines 11 & 17, Error: (a) unable take, (b) to try to solving

Sample sentence 1	They are unable take any measures for a better future.
Correction 1	They are unable to take any measures for a better future.
Sample sentence 2	They are experimenting new ways to use the internet and other new
	technology to try to solving some of their problems.
Correction 2a	They are experimenting new ways to use the internet and other new
	technology to solve some of their problems.
Correction 2b	They are experimenting new ways to use the internet and other new
	technology for solving some of their problems.
Correction 2c	They are experimenting with new ways to use the internet and other new
	technology for resolving some of their problems.

10 Auxiliaries (do/did), lines 7 & 8, Errors: (a) were not understand, (b) countries not match.

Sample sentence I	They not understand much about African tribal traditions.
Correction 1a	They did not understand much about African tribal traditions.
Correction 1b	They were not informed about African tribal traditions.
Correction 1c	They were not aware of African tribal traditions.
Correction 1d	They are not able to understand African tribal traditions.
Sample sentence 2	They did not understand much about African tribal traditions,
	why the borders of these countries not match the traditional borders.
Correction 2a	They did not understand much about African tribal traditions, that's why the
	borders of these countries do not match the traditional borders
Correction 2b	They did not understand much about African tribal traditions, that's why the
	borders of these countries did not match the traditional borders.

11 Word order, lines 9 & 13, Error: (a) problems facing Africa today, (b) rapidly has spread

Sample sentence 1	Many of the problems Africa facing today have been worsened by fighting over control of the government.
Correction 1a	Many of the problems facing Africa have been worsened by fighting over control of the government.
Correction 1b	Many of the problems facing Africa today have been worsened by fighting over control of the government.
Correction 1c	Many of the problems Africa faces today have been worsened by fighting over control of the government.
Correction 1d	: Many of the problems, which Africa is facing today, have been worsened by control of the government.
Correction 1e	Many of the problems Africa is facing today have been worsened by fighting over control of the government.
Correction 1f	Many of the problems facing Africa nowadays have been worsened by fighting over control of the government.

	Correction 1g	Many of the problems facing Africa nowadays have been worsened by
		fighting over control of the government.
	Sample sentence 2	HIV, the virus that causes AIDS, rapidly has spread in Africa.
	Correction 2a	HIV, the virus that causes AIDS, has rapidly spread in Africa.
	Correction 2b	HIV, the virus that causes AIDS, has spread rapidly in Africa.
	Correction 2c	HIV, the virus that causes AIDS, has spread in Africa rapidly.
	Correction 2d	HIV, the virus that causes AIDS, is rapidly spreading in Africa.
12	Auverb, mies 4 &	14, Errors: (a) changed dramatic, (b) spread quick
	Sample sentence 1	It is because the situation has changed dramatic.
	Correction 1a	<u> </u>
	Correction ra	It is because the situation has changed dramatically .
	Correction 1b	It is because the situation has changed dramatically . It is because of a dramatic change in situation.
		· · · · · · · · · · · · · · · · · · ·
	Correction 1b	It is because of a dramatic change in situation.
	Correction 1b Sample sentence 2	It is because of a dramatic change in situation. Other diseases have also spread quick .