



2018-12-01

Promoting Second Language Learning Through Oral Asynchronous Computer-Mediated Communication

Eric H. Young
Brigham Young University

Follow this and additional works at: <https://scholarsarchive.byu.edu/etd>

 Part of the [Educational Psychology Commons](#)

BYU ScholarsArchive Citation

Young, Eric H., "Promoting Second Language Learning Through Oral Asynchronous Computer-Mediated Communication" (2018).
All Theses and Dissertations. 7051.
<https://scholarsarchive.byu.edu/etd/7051>

This Dissertation is brought to you for free and open access by BYU ScholarsArchive. It has been accepted for inclusion in All Theses and Dissertations by an authorized administrator of BYU ScholarsArchive. For more information, please contact scholarsarchive@byu.edu, ellen_amatangelo@byu.edu.

Promoting Second Language Learning Through Oral Asynchronous Computer-
Mediated Communication

Eric H. Young

A dissertation submitted to the faculty of
Brigham Young University
in partial fulfillment of the requirements for the degree of
Doctor of Philosophy

Charles R. Graham, Chair
Richard E. West
Troy L. Cox
R. Kirk Belnap
Ross Larsen

Department of Instructional Psychology & Technology
Brigham Young University

Copyright © 2018 Eric H. Young

All Rights Reserved

ABSTRACT

Promoting Second Language Learning Through Oral Asynchronous Computer-Mediated Communication

Eric H. Young

Department of Instructional Psychology & Technology, BYU

Doctor of Philosophy

Learning to speak a foreign language (L2) can be a challenging feat, made all the more challenging when done in only 50 minute, daily increments in class. Oral asynchronous computer-mediated communication (ACMC) provides learners with opportunities to practice spoken communication and evaluate their practice outside the classroom. In this dissertation, I explore methods for classroom integration of oral ACMC, linguistic traits developed in previous oral ACMC studies, methods for determining the effectiveness of oral ACMC, learner beliefs about the effectiveness of oral ACMC activities, and the effects of learners' deliberate practice in a series of oral ACMC activities on 3 measures of L2 fluency. In my first article, a literature review, I found that most studies on this topic focus on the linguistic traits of accuracy, fluency, and pronunciation, and determine L2 growth from oral ACMC activities through learner perceptions of L2 growth not relying on objective measures. In my second article, I analyzed the fluency change of learners who participated in a series of video recording and feedback activities. I found that, although there were few significant results, the activities may be of some benefit to learners in improving their spoken fluency. I also found that structural equational modelling may be of more value for researching classroom-based activities than *t* tests and regression models. In my third article, I investigated the experiences of several learners who participated in the video recording activities described in article two. Based on these learner experiences, I provided key considerations for designing asynchronous video recording assignments. The three articles included in this dissertation will be valuable in highlighting key factors related to the design, development, research, and effective use of oral ACMC activities in foreign language classrooms.

Keywords: asynchronous computer-mediated communication, language learning, video, critical incident technique

ACKNOWLEDGMENTS

This dissertation would not have been possible without the training and support of a number of people. Key among these people has been my wife, Dr. Trisha Young. She has mentored me since I began work on my MA and has continued doing so while pursuing my Ph.D. Beyond her academic help have been the professional and personal sacrifices she has made so that I could begin this round of education.

I would also like to thank Michael Griffiths, who sparked the idea of this dissertation in my mind. During a department seminar, he presented on the use of asynchronous video to communicate with students and create a stronger sense of community in distance and blended learning environments. Since that presentation, I began exploring different applications in foreign language learning contexts.

Each of my committee members has also been instrumental in the success of this dissertation. Charles Graham provided his blended learning expertise as well as his encouragement to stay focused, motivated, and action-oriented, helping me get things done when they needed to be. Rick West has been constantly available to help me test the logic of my arguments and ensure that everything is comprehensible, in addition to lending his qualitative research expertise. Ross Larsen was not only a statistical saving grace but also took time in our meetings to check up on my mental health, a component of the graduate studies experience that can sometimes be overlooked. Troy Cox not only helped me find a suitable means of measuring fluency, but has also helped me keep looking beyond the dissertation so that graduation would not be an endpoint but rather a continuation of my work and studies. Kirk Belnap has been very patient as I have tried to incorporate my designs into his Arabic program. In fact, it is thanks to him allowing me to explore different teaching practices that I was able to try out different uses of

asynchronous video in Arabic classrooms. He was also key in helping me review and apply the notions of deliberate practice into this study.

Many others have helped me along the way, whether preparing me in the past or dedicating hours recently in helping me with analysis, coding, and editing. Among them are Esther Michela, Harold Hendricks, Peter Rich, Randy Davies, Stirling Miller, and Ben Nielsen.

TABLE OF CONTENTS

TITLE PAGE	i
ABSTRACT.....	ii
ACKNOWLEDGMENTS	iii
TABLE OF CONTENTS.....	v
LIST OF TABLES.....	x
Article 1 Tables.....	x
Article 2 Tables.....	x
Article 3 Tables.....	xi
LIST OF FIGURES	xii
Article 1 Figures	xii
Article 2 Figures	xii
Article 3 Figures	xii
DESCRIPTION OF RESEARCH AGENDA AND DISSERTATION STRUCTURE	xiii
SPEAKING PRACTICE OUTSIDE THE CLASSROOM: A LITERATURE	
REVIEW OF ASYNCHRONOUS MULTIMEDIA-BASED ORAL	
COMMUNICATION IN LANGUAGE LEARNING.....	
Abstract.....	2
Introduction.....	3
Methodology.....	9
Inclusion / Exclusion Criteria	9
Relevance.....	9
Outlet type.....	10

Research type.....	10
Examples of inclusion/exclusion.....	10
Search Results.....	12
Using AMOC in Language Learning.....	12
Using AMOC to Develop Specific Language Traits.....	12
Accuracy.....	12
Fluency.....	13
Pronunciation.....	14
Conclusions about these traits.....	15
Methods and Challenges in Using AMOC.....	15
Technological challenges and training.....	16
Preparatory activities.....	20
Project-based learning.....	21
Self-evaluation and revision.....	22
Conclusions regarding AMOC methods and challenges.....	24
Methodologies for Measuring and Analyzing Language Gains in AMOC.....	25
Data Sources.....	26
Data Analyses.....	28
Conclusions on Methodologies.....	29
Conclusions.....	30
References.....	33
Appendix A: Articles Reviewed in this Study.....	39
Appendix B: Comparison of Articles Reviewed.....	42

DELIBERATE PRACTICE OF L2 FLUENCY WITH SELF AND

EXPERT FEEDBACK	43
Abstract	44
Introduction.....	45
Literature Review.....	46
Fluency.....	46
Deliberate Practice	47
Asynchronous Videos in Deliberate Fluency Practice.....	48
Methodology.....	50
Participants.....	51
Task.....	51
Grouping.	53
Expert evaluation.	54
Self-evaluation.	55
Measurement.....	55
Analysis.....	57
Results.....	59
Fluency Change from First to Second Drafts	59
The Effect of Self- and Expert Evaluation on Fluency.....	64
Effectiveness of Order of Treatment.....	71
Discussion	72
Fluency Change from First to Second Drafts	73
The Effect of Self- and Expert Evaluations on Fluency	74

Effectiveness of Order of Treatment.....	75
Limitations	75
Implications.....	77
References.....	79
Appendix A: IRB-Approved Consent Form.....	83
Appendix B: Demographic Information Survey.....	84
KEY INFLUENCES OF LEARNER-CREATED ASYNCHRONOUS VIDEOS	
ON GAINING L2 FLUENCY	85
Abstract.....	86
Introduction.....	87
Critical Incidents and User Feedback	88
Learning Task	90
Methods.....	92
Critical Incident Technique.....	95
Interview recording	95
Interview fidelity.....	95
Exhaustiveness.....	95
Participation rates.....	96
Independent extraction of incidents.....	97
Independent coding	98
Crosschecking	99
Expert opinions	99
Theoretical agreement.....	100

Findings.....	100
Interviews.....	100
Nature of the recording assignments.....	100
Components of the recording assignments.	101
Training.....	106
Surveys.....	106
Rankings.	106
Follow-up feedback.	108
Conclusions.....	110
Determine the Focus, Content, and Format of Feedback Given to the Learners.....	110
Provide Learners with Clear Linguistic Models.....	111
Provide a Variety of Training Types at Multiple Points.....	112
Set up the Assignments in Ways that Encourage Practice.....	112
Create Realistic, Engaging Prompts that Align Closely with Course Content.....	112
References.....	114
Appendix: Categorized and More Inclusive List of Critical Incidents.....	117
DISSERTATION CONCLUSION.....	119

LIST OF TABLES

Article 1 Tables

Table 1	<i>Comparison of Asynchronous and Synchronous Communication</i>	6
Table 2	<i>Summary of Inclusion/Exclusion Criteria</i>	9
Table 3	<i>Examples and Non-examples of Articles Found in the Literature Search</i>	11
Table 4	<i>Frequency of Data Types</i>	26
Table 5	<i>Frequency of Data Analysis Types</i>	26

Article 2 Tables

Table 1	<i>Learner Grouping and Evaluation Types</i>	54
Table 2	<i>Quotients of Skewness and Standard Error of Skewness as an Assessment of Distribution</i>	60
Table 3	<i>Results of Paired-Samples t Tests for nMSD</i>	61
Table 4	<i>Results of Paired-Samples t Tests for nMLUsec</i>	62
Table 5	<i>Results of Paired-Samples t Tests for nMLU syll</i>	63
Table 6	<i>Results of Paired-Samples t Tests for nASPD</i>	64
Table 7	<i>Results of Independent Samples t Tests for the Change in Fluency for the First Treatment Period</i>	65
Table 8	<i>Regression Equations for Change in Normalized Fluency Measures from First to Last Video of First Treatment Period Covaried by Completion Rate</i>	65
Table 9	<i>ANOVA F-test for the Regression Equations for Change in Normalized Fluency Measures from First to Last Video of First Treatment Period Covaried by Age, Year in School, Major of Study, Treatment Group, Class, Gender, Previous Arabic Experience, and Previous Non-Arabic Language Experience</i>	66

Table 10	<i>Results of LGMs Exploring the Effect of Treatment Type on Fluency for Videos 1–3</i>	67
Table 11	<i>Effects of Completion Rate on the Slopes of the LGMs on Their Respective Outcomes</i>	69
Table 12	<i>Results of an LGM Exploring the Effect of Completion Rate on Fluency for Videos 3–6</i>	70
Table 13	<i>Results of LGMs Tracking Unconditional Fluency-Change</i>	71
Table 14	<i>Results of an LGM Comparing the Effect of the Order of Treatment from Videos 1–6 on the Slope of the LGM</i>	71
Table 15	<i>Results of an LGM Comparing the Effects of Order of Treatment for Videos 3–6 on the Slope</i>	72

Article 3 Tables

Table 1	<i>Critical Incident Categories and the Percentage of Interviewees Mentioning Each Category</i>	97
Table 2	<i>Rankings of Top Five Critical Incidents, Influences, and Aspects of Video Recordings</i>	107

LIST OF FIGURES

Article 1 Figures

<i>Figure 1.</i> Relationship of online and multimedia-based communication to speaking proficiency	4
--	---

Article 2 Figures

<i>Figure 1.</i> Timing of intervention activities. Video assignments are assigned and submitted on the week they fall next to.....	52
<i>Figure 2.</i> Plot for the LGM of normalized ASPD based on completion rate.	69

Article 3 Figures

<i>Figure 1.</i> GoReact interface where learners mark (V)ocabulary/pronunciation, (P)auses, and (R)epairs and write comments.	92
---	----

DESCRIPTION OF RESEARCH AGENDA AND DISSERTATION STRUCTURE

This journal-ready multiple article format dissertation explores the topic of oral asynchronous computer-mediated communication (ACMC) in foreign language learning environments. Given the necessity of practice to develop spoken language skills and the challenge in finding enough class time for that practice, oral ACMC provides an avenue for extending language practice time outside the classroom. In this dissertation, I present three studies, showing the current state of oral ACMC research, examining the pedagogical effectiveness of oral ACMC-based learning activities, and recommending critical factors in designing future oral ACMC activities.

In the first article, “Effective use of oral asynchronous computer-mediated communication in language learning,” I present a literature survey of recent oral ACMC studies. I outlined several ways oral ACMC has been used to promote language learning in terms of key linguistic features. Additionally, I presented the most common methods for studying oral ACMC and determining the effectiveness of its use in language learning environments. Appendix A lists the references of all the research reviewed. Appendix B contains a table showing the different types of data collected for each article considered in this literature review and the methods used to analyze the data. This article was published by *EuroCALL Review* in March 2018 and was formatted according to this journal’s style and requirements.

In the second article, “Comparing the effect of self-evaluation and expert evaluation on the development of fluency using oral asynchronous computer-mediated communication,” I analyzed the effect of learner video draft creation and feedback types on second language fluency gains. Appendix A contains the IRB-approved consent form used in this study. Appendix B contains the survey used to help determine the a priori matching. This study would

be a good fit for the following three journals: (a) *Language Learning and Technology* because of its focus on both foreign language learning and technology, and because of its metrics (1 internal reviewer, 3 external reviewers; impact factor = 2.113 for 2017), and its international prestige; (b) *ReCALL* for the same reasons listed above, and because of its metrics (2 external reviewers ;impact factor = 2.206 for 2017) and its international prestige; and (c) *Innovation in Language Learning and Teaching* because of its focus.

In the third article, “Key influences of learner-created asynchronous videos on gaining L2 fluency,” I explore the critical incidents, influences, and aspects of the learner video creation intervention to be used in article two to find what learners believe led to either their success or challenges. The appendix contains additional critical incidents not used in the final analysis. This study will be a good fit for the following three journals: (a) *IALLT Journal of Language Learning Technologies* because of its acceptance of qualitative studies related to technology use in higher education, and because of the importance of the sponsoring institution in language learning technologies; (b) *Language Learning & Technology* (1 internal reviewer, 3 external reviewers; impact factor = 2.113 for 2017); (c) *ReCALL* (2 external reviewers; impact factor = 2.206 for 2017).

SPEAKING PRACTICE OUTSIDE THE CLASSROOM: A LITERATURE REVIEW OF
ASYNCHRONOUS MULTIMEDIA-BASED ORAL COMMUNICATION
IN LANGUAGE LEARNING

Young, E. H., & West, R. E. (2018). Speaking practice outside the classroom: A literature review of asynchronous multimedia-based oral communication in language learning. *EuroCALL Review*, 26(1), 59–78. doi.org/10.4995/eurocall.2018.8599

Speaking Practice Outside the Classroom: A Literature Review of Asynchronous
Multimedia-based Oral Communication in Language Learning

Abstract

Classroom instruction provides a limited amount of quality speaking practice for language learners. Asynchronous multimedia-based oral communication is one way to provide learners with quality speaking practice outside of class. Asynchronous multimedia-based oral communication helps learners develop presentational speaking skills and raise their linguistic self-awareness. Twenty-two peer-reviewed journal articles studying the use of asynchronous multimedia-based oral communication in language learning were reviewed (a) to explore how asynchronous oral communication has been used to improve learner speaking skills, and (b) to investigate what methodologies are commonly used to measure and analyze language gains from using asynchronous multimedia-based oral communication to improve learner speaking skills. In this study we present three principal findings from the literature. First, asynchronous multimedia-based oral communication has been used in conjunction with a variety of instructional methods to promote language gains in terms of fluency, accuracy and pronunciation. Second, the methods found in this review were technical training, preparatory activities, project-based learning, and self-evaluation with revision activities. Third, the majority of previous studies demonstrating the effectiveness of these methods have relied on learner perceptions of language gains rather than on recordings of learner speech.

Keywords: oral, online, asynchronous, video, audio, language learning

Introduction

In order for foreign language learners to succeed, they need a large quantity of high-quality language practice. Although Clifford described time on task, or quantity, as “the primary determiner of language acquisition” (2002), it has also been described as “a necessary, but not sufficient, condition for learning” (Karweit, 1984, p. 33). Hirotani and Lyddon (2013) argued that quality of practice, exemplified in their study by an awareness-raising activity, is an important factor in the language learning.

Media-based oral communication can increase the quantity and improve the quality of language practice by providing more opportunities for speaking and more opportunities to raise learner awareness. Multimedia-based oral communication includes a variety of communication types, such as video conferencing through Skype, posting vlogs on YouTube, and turn-based video conversations using a voiceboard. Lin (2015) lauded the affordances of oral computer-mediated communication (CMC: an important type of multimedia-based oral communication) in his meta-analysis, stating that the “features of CMC seem to provide opportunities to create a social interaction context with more flexibility that cannot be afforded in a traditional face-to-face environment” (p. 262). Here it is useful to recall Clark’s (1994) criticism of many media-related studies, that media itself does not influence learning. Rather it is the instructional method that influences learning. Referring to his previous studies, Clark summarized his argument, stating, “any necessary teaching method could be designed into a variety of media presentations (p. 22). On the other hand, however, it is important to note that certain media and technologies provide *affordances* that may not be otherwise available or that are more effectively used with those media and technologies.

In his book on distance and blended (a.k.a., hybrid) learning, Graham (2006) stated that online learning environments provide learners with flexibility in communicating outside the classroom. By communicating online, learners may increase their opportunities for speaking practice. Additionally, the digital nature of online communication makes it easier for learners to record and review their speech, allowing them to develop linguistic self-awareness. Both the opportunities and self-awareness promote increased speaking proficiency. Figure 1 illustrates these affordances and their relationship.

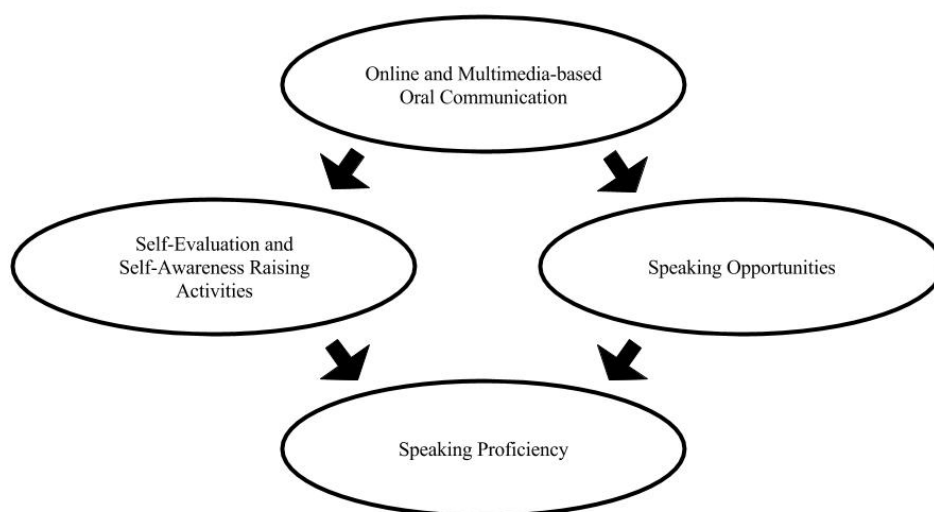


Figure 1. Relationship of online and multimedia-based communication to speaking proficiency

Lin (2015) discussed these affordances in his meta-analysis of CMC use. Although he referred specifically to text-based communication, the affordances also apply to oral communication. He stated that CMC “provides L2 learners with an environment to practice language production at a reduced rate. The relatively reduced rate of exchange and lag-time induced by the text-chat software allows L2 learners ‘more time to both process incoming messages and produce and monitor their output’ (Sauro & Smith, 2010, p. 557)” (Lin, 2015, p. 264).

Similarly, in her meta-analysis of 14 studies involving CMC, Ziegler (2016) argued that CMC use provides learners with an opportunity to “notice [the] gaps between their interlanguage and the target language” (p. 575). Because of the time lag that Lin (2015) referred to, Ziegler (2016) found that CMC may be more beneficial to language learning than face to face communication in the target language in terms of developing productive language skills. So, although online oral activities may make use of the same methods that face-to-face activities use, the affordances of online activities may make them at least as effective as, and sometimes more practical than, face-to-face activities by increasing the quantity and quality of oral language practice.

Communication can be categorized as either synchronous, having little or no lag time, or asynchronous, having a long lag time, based on Graham’s (2006) description of distance learning environments (see Table 1). Although asynchronous and synchronous communication are similar in some ways, asynchronous communication provides opportunities that synchronous communication (or even classroom speaking activities) does not. First, synchronous communication is more conducive to interpersonal speaking. Ziegler (2016), in her synthesis of synchronous computer-mediated communication (SCMC) use, situated SCMC within the interaction hypothesis, arguing that it provides opportunities for interaction and negotiation of meaning. Asynchronous oral communication, on the other hand, can be considered a type of presentational speaking, a necessary skill in many occupations—see the American Council on the Teaching of Foreign Language’s (2012) description of modes of communication for more information. However, it could be argued that even synchronous conversations consist, to a degree, of a series of mini-presentations. Whereas Kitade (2000) rightly argued that interlocutors need interaction skills and pragmatic competence when responding to one another

in synchronous conversations, they sometimes do so by providing complete, continuous responses or by sharing anecdotes.

Table 1

Comparison of Asynchronous and Synchronous Communication

Asynchronous	Synchronous
Prepared	Spontaneous
Targets presentational speaking	Targets interpersonal speaking
Disposed to formal evaluation	Disposed to impromptu, informal evaluation
Revisionary	Single occurrence

Second, asynchronous communication more naturally promotes planning before the speech act whereas synchronous communication tends to be more spontaneous. Crookes (1989) discussed the value of pretask planning to improve non-spontaneous language output. In his study, 40 Japanese learners of English participated in two oral explanation tasks. Group 1 (n=20) was given no preparation and planning time before participating in the task. Group 2 (n=20) was given 10 minutes of preparation and planning before the tasks. Crookes found that learners who planned their output generally produced a greater variety of lexis, more complex language, and more detailed descriptions.

Third, asynchronous communication more naturally allows learners to watch or listen to their own performance and conduct self-evaluation. Instructors and learners in many domains have used video recordings of learner behavior to increase self-awareness and determine what skills they need to focus on. Examples can be found in sports (Hastie, Brock, Mowling, & Eiler, 2012) and medicine (Jamshidi, LeMasters, Eisenberg, Duh, & Curet, 2009). In Jamshidi et al.'s (2009) study involving junior surgeons practicing laparoscopic suturing skills, learners benefited

from reviewing video recordings of their practice attempts. The learners grew in terms of both self-awareness and skill in part because video recording “provides a matrix of information identical to what was available during the operation itself” (p. 625). This is particularly important in language learning, where the learner’s memory is taxed while trying to create a message to the point that they may not be wholly aware of the actual language they are producing. Video provides them with the opportunity to hear exactly what they said. In fact, Jamshidi et al. (2009) argued that this type of video review can not only be used as for post-performance assessment but also in pre-performance planning (p. 625).

Fourth, because of its recorded nature, asynchronous communication enables learners to revise and rerecord their performance so that they can publish their best version. Learners have long had the opportunity to improve their composition writing by creating several drafts before submitting a final version. Although, learners can also practice oral presentations before a live audience (e.g., a classmate) or in front of a mirror prior to their final performance, this asynchronous multimedia-based communication (AMOC) provides another outlet for this kind of practice that can be done on the learner’s own time. Another benefit that live practice does not afford, however, is that AMOC allows the learner to select the best video or audio draft to submit, rather than having to submit the final performance. Additionally, in some draft-writing processes, learners are even asked to focus on revising a specific element of their writing (e.g., spelling or paragraph structure). Castañeda and Rodríguez-Gonzalez (2011) incorporated this kind of process in their study of nine university-level learners of Spanish and found that learners increased in terms of speaking, analytic, and evaluation skills.

Although AMOC is generally better suited to promoting self-awareness, revision, and presentational speaking skills, synchronous communication seems to be the more popular of the

two in blended language learning environments. It may be easy to think that synchronous communication is better for improving learner speaking proficiency, given its shorter lag time and better simulation of face-to-face conversations. Because of this, we risk falling into the trap of relegating AMOC to the status of technologies we only use if we do not have bandwidth and hardware that supports synchronous conversation. Yet, given that AMOC provides different affordances than what synchronous communication offers, asynchronous communication can serve different purposes than synchronous communication.

However, even though AMOC *can* provide learners with opportunities to develop their linguistic self-awareness and improve their speaking skills, there is no guarantee that learners *will* make these gains by participating in oral asynchronous activities. The purpose of this literature review, then, is to explore how AMOC has been used to improve speaking skills. Additionally, we examine the methodologies that previous research has used to measure improvements in speaking skills. Thus, in this study we will address the following research questions:

Question 1: What language traits are being promoted with AMOC?

Question 2: What are the challenges to effective use of AMOC?

Question 3: What methods and activities have been used in conjunction with AMOC?

Question 4: What methodologies are commonly used to measure and analyze language gains from using asynchronous multimedia-based oral communication to improve learner speaking skills?

Methodology

Literature was located using Academic Search Premier, ERIC, JSTOR, and Scopus. The following combinations of search terms were used: *asynchronous video + language*, *asynchronous CMC + language*, *asynchronous + speaking + language*, *video-mediated communication + language*, *vlog + language*, *Wimba + language*, *oral CMC*, *video drafts + language*, and *blended learning + video + language*. Literature was limited to that published before early 2016.

Inclusion / Exclusion Criteria

The following criteria were used to determine which studies to include in this analysis. They are relevance, outlet type, and analysis methods (see Table 2).

Table 2

Summary of Inclusion/Exclusion Criteria

Criterion	Definition
Relevance	University level learner-created oral asynchronous audio or video productions; research focuses on language gains
Outlet type	Peer reviewed journal articles
Analysis methods	Qualitative and quantitative methods

Relevance. We used the following criteria to determine if studies were sufficiently relevant to this discussion:

- Studies must address asynchronous audio or video communication.
- Videos must be learner created.
- Studies must discuss how learners improved language skills by producing videos.

- Studies must discuss university level class implementation in order to maintain comparability between studies.

Outlet type. Only peer-reviewed journal articles were included in this review. Book chapters and conference proceedings were not included. Conference proceedings, although useful, were not included in order to maintain a higher standard for inclusion in this literature review.

Research type. Only articles including qualitative and quantitative studies were included. This criterion is particularly relevant for research question 1 where both empirical and qualitative information clarified how well learning is taking place. For instance, in Kormos and Dénes' (2004) study, speaking fluency was described in terms of specific, empirical measurements, which enables us to compare fluency across studies. On the other hand, Castañeda and Rodríguez-González (2011) shared learner feedback from self-evaluations after participating in an asynchronous video intervention. While this qualitative data did not provide a clear means of comparing learning effectiveness as did Kormos and Dénes' (2004) study, it did provide insights into the learners' experiences, and it provided other information that might not have been solicited or considered in an empirical study. For instance, one learner discussed the concept of anxiety in their responses (2004), which is an important aspect of the use of asynchronous video communication but would not necessarily be considered in a comparison of fluency gains. Theory and design articles were not included unless they also included either a qualitative or quantitative study showing the effect of their theory or design in practice.

Examples of inclusion/exclusion. Table 3 displays examples of articles found during the literature search along with an indication of whether the example article met a given criterion ("X") or did not meet the criterion ("—"). This is meant to give an explanation of our decision

process in choosing which articles to include for review. Of the examples shown in Table 3, only Hirotani and Lyddon (2013) met all three criteria and was, therefore, the only one included in this literature review. Tiraboschi and Iovino (2009) presented activities and a related technology but did not focus on the learning effects of implementing the activities and technology or present any data. Hirotani's (2009) article focused on text-based CMC rather than audio or video CMC. Ono, Onishi, Ishihara, and Yamashiro (2015) presented a paper that was published in the conference proceedings, which did not meet the requirement of being a peer-reviewed journal article. Lamy and Goodfellow (1999) focused on text-based CMC, but also focused on language used during ACMC tasks, rather than language gained from using the tasks.

Table 3

Examples and Non-examples of Articles Found in the Literature Search

Example/Non-example	Relevance	Outlet Type	Analysis	
			Methods	Reason for Exclusion
Tiraboschi & Iovino (2009)	—	X	—	No data/design showcase
Hirotani (2009)	—	X	X	Text-based CMC
Hirotani & Lyddon (2013)	X	X	X	NA
Ono, Onishi, Ishihara, & Yamashiro. (2015)	X	—	X	Conference proceeding in book
Lamy & Goodfellow (1999)	—	X	X	Text-based CMC; does not focus on language gains

Search Results

Using the aforementioned search terms and inclusion/exclusion criteria, 22 articles were located (see citations for these articles in Appendix A).

Using AMOC in Language Learning

From this pool of articles, we identified several factors that affect the effectiveness of AMOC activities in language learning contexts. This section begins with a description of the linguistic traits that AMOC activities has been used to improve, then moves to a discussion of challenges inherent in using AMOC, and then concludes with a discussion of the effectiveness of various methods of using AMOC to improve the linguistic traits that will be described.

Using AMOC to Develop Specific Language Traits

In this section, we address the question of *what language traits are being promoted with AMOC*. We will focus on accuracy, fluency, and pronunciation. Although AMOC is used to help learners develop several different linguistic traits, we found that these particular traits need to be treated with more rigor.

Accuracy. By using AMOC, learners are able to increase the accuracy of their speech. In a study on the effects of using AMOC in an ESL writing course, Engin (2014) interviewed participants and analyzed questionnaires, finding that students believed their linguistic accuracy increased as a result of creating their videos. Learners were expected to create English writing explanations (tutorials) for other students in their class in video format. Because of the responsibility of teaching placed upon them and peer dependence on their creating a clear, effective explanation, learners felt compelled to produce linguistically accurate explanations and reduce the number of mistakes in their performance. Engin cited one learner's interview response that the video activity helped their accuracy: "It is a good thing to worry about our

English because we improve our English” (2014, p. 19). Unfortunately, it is not clear in what ways learner speech increased in accuracy nor the learner’s basis for determining whether they increased in accuracy or not. Although Engin’s findings suggest that AMOC can be used to improve accuracy, additional data and analysis procedures would provide a more rigorous, reliable and trustworthy basis for determining that learner speech became more accurate through producing these videos.

Fluency. Learners using AMOC are also able to develop fluency. In his study of Japanese EFL students, Gromik (2012) found that learners increased their speech rate by 37% over the course of a 13-week video production intervention, comparing average speech production of the first and final weeks. Although the average speech rate of the first week was significantly lower than all subsequent weeks, suggesting that some of the learners’ improvement may be attributed to familiarization with the task and the technology, Gromik demonstrated a general increase in speech rate attributable to learner production of asynchronous videos.

Despite the generally positive findings of Gromik’s (2012) study, his study leaves us with several questions. For instance, Gromik only considered the speech rate of short videos, where the task limited learners to 30-second video clips. It is unclear whether the learners in this study could sustain this speech rate. It is also unclear whether producing longer videos would offer the same advantage in helping learners develop a higher peak speech rate or a higher consistent speech rate. Gromik also considered only two closely related aspects of fluency: number of words produced and speech rate, or number of words produced per second.

While Gromik’s (2012) inclusion of two fluency measures is valuable, it does not represent the wide array of fluency measures available to researchers. In their study on the relationship between proficiency and fluency, Baker-Smemoe, Dewey, Bown, and Martinsen

(2014) presented three major categories of speech fluency, each characterized by several different aspects, based on Segalowitz's (2010) work on fluency. These categories are cognitive fluency, perceived fluency, and utterance fluency. Cognitive fluency refers to the ease with which a speaker is able to create and produce speech; perceived fluency refers to native speaker judgments of how easily the learner produces speech; and utterance fluency refers to measurable aspects of learner speech, including speech rate, hesitations and pausing.

Although Gromik's (2012) study demonstrated the potential value of using AMOC to improve learner fluency, more evidence is needed in order to generalize his findings. Further research should consider the various categories of fluency and the effect of AMOC on fluency in longer videos.

Pronunciation. AMOC has also been shown to help learners develop their pronunciation. In a study involving 39 students of French, Lepore (2014) linked AMOC participation to the learners' perceptions of improvement in their pronunciation. Learners in this study used VoiceThread to produce three audio recordings in response to instructor-created prompts and then commented on one another's recordings. After submitting their recordings, learners completed self-assessments, rating their pronunciation during the recordings.

As with Engin's (2014) findings on increased accuracy, relying solely on the perceptions of untrained learners in Lepore's (2014) study renders the validity of the findings questionable. Although Lepore's self-assessment form provides multiple questions to help the learners think about their pronunciation development (e.g., pronunciation compared to peers' pronunciation, pronunciation improvements as a result of using VoiceThread, and accuracy of specific vowel and consonants in French), it neither provides clear guidance in rating their pronunciation nor provides guidance on what should be rated. In this case, a rubric identifying front rounded

vowels, front unrounded vowels, back vowels, and difficult French consonants (e.g. /ʁ/) along with a rating scale, a series of descriptions of performance (e.g. native-like, somewhat native-like), or a series of characteristics (e.g. vowel was not rounded but was at correct height) might guide learners to more accurately and reliably assess their own pronunciation, as well as guide them to improving their pronunciation.

Conclusions about these traits. AMOC has been used to promote language gains in terms of accuracy, fluency, and pronunciation. However, it is not clear what aspects of accuracy were improved through AMOC. For instance, it may be that oral ACMC activities are conducive to lexical accuracy but not syntactic accuracy, or the converse. Fluency seems more clearly affected by AMOC activities, as studies have used more clear and varied measurements to determine fluency gains. Finally, although AMOC was shown to promote pronunciation gains, the evidence supporting this notion is insufficient. This may be remedied through the use of more rigorously developed self-rating systems, through native-speaker raters, or through acoustic measurements, such as comparing learner consonant production with native-speaker production using PRAAT, a popular phonetic analysis program. In summary, AMOC has been shown to have the potential to promote language gains in various linguistic aspects, but additional studies and more rigorous research methods are needed to confirm this.

Methods and Challenges in Using AMOC

Although AMOC has been shown to be a promising medium for helping learners increase their fluency, accuracy, and pronunciation, the mere inclusion of AMOC in a learning environment does not guarantee these increases. The question remains, then, of how to effectively incorporate AMOC into a course curriculum and how to deal with the challenges that inevitably arise. In this section, we address research question 2 by discussing technological

challenges that have arisen in previous studies, and address research question 3 by discussing methods and activities that have contributed to the effective use of AMOC in language learning. The methods and activities discussed are training activities, preparatory activities, project-based learning, and self-evaluation combined with revision.

Technological challenges and training. Although many factors affect the quantity and quality of language learning experiences, whether in a classroom or online, technological challenges, in particular, affect the learning experience during AMOC activities. A variety of technological challenges exist. Poor internet connection is a common challenge that can be experienced in any location. In their study on Malaysian learners using both audio and video recordings, Bakar, Latiff, and Hamat (2013) reported that even learners at a university experienced connectivity problems, affecting their access to the AMOC activities and thereby their level of participation. Hung's (2012) learners in Taiwan also experienced poor internet.

In addition to internet problems, learners may experience hardware deficiencies and malfunctions. Learners in Bakar, Latiff, and Hamat's (2013) study experienced hardware malfunctions that made it impossible to record their voice. Gleason and Suvorov (2012) stated that their learners also had trouble saving and editing their recordings. In Gromik's (2012) study, some learners were unable to upload video files because they were too large. As these video recordings were 30 seconds or shorter, it seems likely that either some learners were unaware of how to select different codecs and file containers for exporting their video or that the recording software they used did not allow them the option to select different codecs or containers. Hung (2012) confirmed this challenge by stating that his learners had difficulties in converting video files into different formats. This was further complicated by the fact that the vlog (video web log) system used in his study only supported a limited set of file formats. Shih

(2010) clarified the problem of file format and file size, adding that internet speed is an important and related factor. Thus, with higher internet speeds, file size may not always be a problem, but with lower internet speeds it will be.

Regarding the problem of access to video recording equipment and editing software, Fukushima (2002) argued that in 2002 the cost of equipment and software licenses was, in fact, not an inhibiting factor for implementing video projects in a language class. By 2016, the affordability and availability of basic editing software and recording equipment has likely increased, leading to better access. This is particularly true when one considers that many university students in the United States own a mobile phone capable of recording high definition videos and performing basic video editing tasks, allowing them to record and edit at any time and in any place. Advanced editing functionality is not necessary for most AMOC tasks, which only require the learner to record a simple video, review it, and then record an additional take rather than splice video segments.

However, because not all learners have mobile phones, or their phones cannot record or edit, it is important to provide other means of recording and editing video files. One way to make recording equipment and editing software available to learners is through university media labs. Some universities offer multimedia labs that loan recording equipment and provide computer stations with editing software. Some even go so far as to offer training in the use of the equipment and software. One drawback to these labs, however, is that they may not provide a suitable environment for recording. As Lepore (2014) stated, a lab setting might lead to some learners reducing their recording quality by speaking softly so as not to disturb other lab users. Background noise might also interfere with recording quality. Despite these drawbacks, labs

offer a possible solution to hardware and software challenges, and both learners and instructors are frequently unaware of their existence at their university.

Compounding the technological challenges, many learners do not have sufficient experience using the hardware or software needed to participate in AMOC. Responding to this lack of experience, Bakar, Latiff, and Hamat (2013, p. 232) stated that their learners would benefit from technical training “so that they are familiar with the online devices and would feel less awkward when utilizing the features of the online tools.” One example of this kind of training took place in Abuseileek and Qatawneh’s (2013) study where learners were provided with basic instruction in using the AMOC software. Similarly, learners in Fukushima’s (2002) study were trained in video and audio editing.

In 2011, Castañeda and Rodríguez-González conducted a study on the effects of self-evaluation and iterative video speech revisions on learners’ linguistics self-awareness and speaking skills. In this study, nine intermediate level Spanish language learners participated in a training activity in which they submitted trial videos prior to participating in the intervention. They created a trial video, following the same procedures they would use to create the videos for the intervention. While the researchers did not mention any specific instruction in how to use the hardware or software, learners nevertheless gained experience in the recording and uploading processes that were required of them in the intervention.

The researchers (Castañeda & Rodríguez-González, 2011) analyzed the learners’ self-evaluation forms to determine if learners felt they had made improvement. In their study, Castañeda and Rodríguez-González did not report any learner dissatisfaction with AMOC caused by technological problems. This may be attributed in part to the carefully organized learning

activities—where learners participated in four cycles of video recordings and subsequent self-evaluation prior to final submission—but also in part to the technical training learners received.

On the other hand, some learners in Dona, Stover, and Broughton's (2014) study who attended a software training session at the beginning of the course still reported having technological challenges. The researchers cited low learner tolerance for learning new technologies as one cause for this problem, and unclear tutorials as a second. While it is not expected that any training activity would solve all technological challenges, a clear description of the training provided would help in discovering how the training could be clearer and how to adapt the training to learners with low tolerance for new technologies.

In Goulah's (2007) ethnographic case study of eight Japanese language learners, learners were not given any formal training on how to use the recording hardware or editing software. Rather, students with prior experience in recording and editing (whether they gained their experience prior to the course or during the first cycle of the intervention activity) became the experts in the second cycle and assisted other learners at that point. In this case, training was done informally by peers, rather than as a formal instructional session by the instructor or researcher. The value in this approach is that learners may, in fact, learn more from someone with a similar status and may learn more because they are receiving instruction while working with the hardware or software. The danger is that instructors cannot guarantee they will have learners with prior experience, and that it may take learners a much longer time to familiarize themselves with the hardware and software before being able to train their peers.

Although it appears training is valuable in alleviating some technological challenges that learners face, there are different ways of providing that training, and it should be carefully designed. Training may be conducted either formally by the instructor or another expert (Dona,

Stover, & Broughton, 2014), or by a more knowledgeable peer (Goulah, 2007). Knowing which learners have prior experience with hardware and software is invaluable if peer-to-peer training is to be expected. Training should also be tailored to the particular learners as much as possible. Many learners are eager to work with new technology, but others are wary of it (Dona, Stover, & Broughton, 2014). Finally, in designing AMOC learning activities, designers must consider learner access to recording hardware and software in the first place. Some may be able to use a mobile phone or personal computer, but others may need access to a lab where they can make their recordings. Yet regardless of the exact nature of the training, training should be provided as many learners lack the skills and equipment necessary to make their recordings, and addressing these deficiencies will help learners to focus on their languaging and not on the technological aspects of the activities.

Preparatory activities. One of the factors that increases the effectiveness of AMOC in developing speaking proficiency is the inclusion of a preparatory activity. Crookes (1989) described planning as a type of preparatory activity in his seminal paper involving 40 Japanese learners of English. He cited “consistent, small- to medium-sized effects in favor of the planned condition” (p. 379), as compared with a control group who did not have planning time. Preparatory activities can take a variety of forms. Bakar, Latiff, and Hamat (2013) described a simple preparatory activity in which learners were given “time to construct and develop their ideas or thoughts” (p. 232) prior to making their audio and video recordings. This preparation enabled the learners to produce more complex ideas. In order to create their video tutorials, Engin’s (2014) learners conducted their own research on their tutorial subjects, finding, evaluating, selecting, and finally summarizing their sources. This task made the learners

responsible for their learning and pushed them to spend time becoming very familiar with it, resulting in students both becoming experts on their topic and developing speaking proficiency.

Goulah (2007) outlined a more complex preparatory activity. Prior to recording their videos, learners in Goulah's study watched videos related to their video topic and then created a storyboard for their video. The storyboard process involved drafting, presenting, negotiating, and finally settling on ideas as a group. Essentially, learners moved from input, to output, and finally to revision of their output, resulting in exposure to authentic language and more time on task. This kind of preparatory activity takes the focus off languaging, as Knouzi, Swain, Lapkin, and Brooks (2010) use the term, for the sake of language and encourages learners to focus on task completion. Learners were able to experience a real need for language and a purposeful interaction in the target language.

Project-based learning. Incorporating AMOC tasks through project-based learning (PBL) can be an effective method of developing learner speaking skills. PBL does this by creating an authentic need to use the target language and by encouraging learners to use a variety of their target language skills and knowledge. In Goulah's (2007) study involving eight intermediate learners of Japanese, learners followed a sequence of project-related activities in which they created commercials responding to challenging political and environmental questions. Their project participation resulted in both an increase of content knowledge and language gains.

Fukushima (2002, p. 353) conducted a study on the effects of PBL in which seven learners collaborated to produce a video promoting Japanese language learning. He described their participation as "self-directed," highlighting that learners assigned their own tasks, set their own schedule, wrote their own scripts, and evaluated and revised their own performance. The

result was that learners produced an authentic linguistic artifact that demonstrated and developed some of their language skills but did not encourage the level of linguistic output and development that the researcher had hoped for. Although language use was considered and reported on, Fukushima focused more attention on motivation and the development of technical skills than on proficiency and performance. A more thorough analysis of the learners' performance in terms of linguistic dimensions, such as accuracy, fluency, and pronunciation, would allow for comparisons with similar learners and allow for a long-term study analyzing the learners' linguistic development.

Although neither Goulah's (2007) nor Fukushima's (2002) studies suggest PBL as an *efficient* means of bringing about language gains, they both demonstrated that PBL has the potential of creating authentic needs for language learning by motivating learners and giving them opportunities to express themselves. Further studies building on Goulah's (2007) and Fukushima's (2002) work should demonstrate ways in which we can efficiently use project-based oral ACMC to create authentic linguistic needs, motivate learners, and bring about significant language gains.

Self-evaluation and revision. In addition to other methods and techniques of incorporating AMOC into learning environments, researchers have found that self-evaluation helps learners achieve language gains. Due to the recorded nature of asynchronous audio and video, learners are not only able to produce spoken output but can listen to their own performance and discover areas of weakness and areas of strength. For instance, most learners in Hung's (2011) study of Chinese learners of English (76%) agreed that participating in creating vlogs helped them reflect on their learning. One learner described the value of the AMOC project in helping them to become aware of their weaknesses and in being able to make

improvement by stating, “I can redo the clips again and again until they looked [sic] satisfactory” (Hung, 2011, p. 742). Lepore (2014) indicated that self-evaluation through AMOC was one of the factors involved in increasing learner willingness to communicate, which itself leads to increased quantity of practice. Dixon and Hondo (2014) reported positive learner impressions of the value of AMOC in making them more aware of their speech production, enabling them to make corrections.

In 2011, Castañeda and Rodríguez-González conducted a study in which nine university-level learners of Spanish produced videos of themselves responding to instructor-generated prompts. Learners in this study responded to a prompt by recording an initial video draft and conducting an evaluation of their draft. They then recorded a second draft and conducted a second self-evaluation. Learners followed this same 2-draft and 2-self-evaluation process, responding to an altered version of the first prompt, although the drafts were labeled as third and fourth drafts. For the self-evaluation, learners watched their recordings, noting mistakes and then recording an improved version.

Learners in Castañeda and Rodríguez-González’s (2011, p. 491) study reported an increase in learner awareness of weaknesses as well as improvements in their grammatical accuracy, pronunciation accuracy, and fluency. Demonstrating increased awareness, one learner stated, “I also noticed my adjective endings weren’t correct.” Another learner commented on the effect of the self-evaluation and revision cycles, “as we do more recordings, the pauses are becoming less frequent.” Castañeda and Rodríguez-González attributed these gains at least in part to the self-evaluation and revision activities.

Of course, incorporating self-evaluation using AMOC does not automatically lead to language gains. Gleason and Suvorov (2012) found that learners were only partially in agreement

($m=3.78$ based on a 5-point scale) that their language skills increased after using AMOC and conducting a self-evaluation. In fact, some learners' perceptions of the value of the intervention actually decreased after participating. In their study, learners recorded three presentations each to share with their peers. They then watched their recordings later to determine if they had made improvements. There is no mention, however, of asking the learners to evaluate their performance and then make changes to their original recording, or to focus on weak areas in subsequent recordings. It seems that learners did not conduct their self-evaluations until after they had completed all their recordings.

Castañeda and Rodríguez-González's (2011) study demonstrated the potential value of combining AMOC with learner self-evaluation and revision cycles. The self-evaluations informed learners of weaknesses and mistakes that learners addressed in subsequent video drafts. Additionally, learners participated in four cycles of self-evaluation and revision. In contrast, learners in Gleason and Suvorov's (2012) study either did not have or did not take the opportunity to improve their recordings based on their self-evaluations. The result was that many did not feel participation in the AMOC activity led to language gains. Thus, while AMOC can be used to create language gains, a structured approach involving both self-evaluation and revision across multiple cycles is more likely to lead to those gains.

Conclusions regarding AMOC methods and challenges. There are a number of things instructors and designers can do to increase the effectiveness of AMOC activities. First, it is important to investigate the learners' hardware and software needs, provide equipment or a lab environment if necessary, and provide training on the creation and sharing of asynchronous audio and video files. If internet speed is a problem, audio might be a more useful option than video, as audio files tend to be much smaller. Second, preparatory activities will improve learner

performance. Preparatory activities range in simplicity from brainstorming ideas before recording to viewing related input and then creating a storyboard. Third, project-based learning in AMOC creates authentic needs for learning and encourages learners to be more self-directed. Finally, cycles of structured self-evaluation followed by revisions may raise learners' linguistic self-awareness and provide them with the opportunity to learn from their heightened awareness.

With those benefits in mind, it is important to note that these methods will not guarantee effective and efficient learning through AMOC. Designers and instructors must incorporate them appropriately, according to the curriculum and the needs of the particular learners. Furthermore, future research is needed to investigate effective methods of incorporating AMOC into a curriculum and to what degree its successful use can be generalized across university-level language learners.

Methodologies for Measuring and Analyzing Language Gains in AMOC

In this section, we address research question 4. The authors of the articles considered in this review used several methods to determine whether AMOC activities brought about learner language gains. In terms of data type, they analyzed surveys, journals, and reflections; learner audio and video recordings; interview transcripts; and researcher observation notes. Table 4 displays the frequency of use for each data type. In terms of data analysis type, researchers used qualitative analysis, descriptive measurements, quantitative comparison, expert evaluation, and correlation. Table 5 displays the number of studies that used each data analysis type. Each data type and analysis type used by a given study were counted individually. Thus, if a study incorporated surveys, interviews, and recordings, as in Shih (2012), the frequency for surveys, interviews, and recordings would each be increased by one. In this way, the total count for data

types and analysis types equaled more than the total number of studies reviewed. Appendix B displays the data and analysis type(s) considered in each study.

Table 4

Frequency of Data Types

Data type	Frequency
Surveys, journals, and reflections	16
Audio & video recordings	12
Interview transcripts	10
Observation notes	2

Table 5

Frequency of Data Analysis Types

Analysis type	Frequency
Qualitative analysis	16
Descriptive measurements	13
Quantitative comparison	6
Expert evaluation	5
Correlation	3
Unknown / unstated	1

Data Sources

Surveys, journals, and reflections was the most common category of data type for determining whether AMOC activities were effective in promoting language gains. Surveys, journals, and reflections were combined into this single category because they contained the

learners' perceptions of their language gains. Many surveys resembled the journals and reflections in that they provided learners with open-ended questions regarding their learning experience, thus increasing the similarity between survey data and journal and reflection data. For instance, Goulah (2007, p. 65) used surveys to discover that participants felt they learned vocabulary and grammar, referring to his surveys simply as "open-ended questionnaires." Others, however, used surveys to collect data on learner opinions of AMOC technology and activities. One example is Hung's (2011, p. 742) survey, which largely focused on learner attitudes based on a five-point scale, "the vlog helped me reflect on my learning in this course," though it contained a question related to learner perceptions of language gains "the vlog helped me organise learning in this course."

Interview data, while the third most common of the four categories, resembled survey, journal, and reflection data, differing only in that interviewers personally elicited learner responses rather than providing them with written questions. Like surveys, interviews focused on learner perceptions of language gains (e.g., Kirkgöz, 2011), as well as attitudes (e.g., Hung, 2011; Yaneske & Oates, 2010). In fact, survey and interview data proved to be similar such that many researchers did not state which themes emerged from survey data and which emerged from interview data.

Audio and video recordings were used as a source of data in roughly one half of the studies considered in this review (n=12). Recordings were either coded for qualitative analysis (n= 6), measured and assigned descriptive statistics (n = 4), or assessed using expert evaluation (n= 4). Three studies used two different analysis types on the recordings (Kormos & Dénes, 2004; Sun, 2012; Sun & Yang, 2015).

Data Analyses

Qualitative analysis was the most common data analysis type found in this study. The term qualitative analysis as used in this study refers to any type of coding and categorizing activities. Conversation analysis and discourse analysis were included in this category.

Descriptive measurement was the second most common analysis type. This term refers to frequency counts, means, and standard deviations. It was frequently used in conjunction with qualitative analysis, as in Shih (2010). In his study, Shih counted the frequency of codes found in learner reflections, and calculated means for survey responses. However, some studies provided empirical descriptions of learner language based on their recordings. For instance, Kormos & Dénes (2004, p. 154) reported 13 statistics, including speech rate, number of words, and mean length of run.

Quantitative comparison refers to quantitative tests used to compare either survey data or learner performance on recordings. In one of the studies (Gromik, 2012), the researcher used a *t*-test to compare learner opinions of the value of using a mobile phone in AMOC activities. In the other five studies using quantitative comparison, the researchers assessed linguistic performance by analyzing recordings and language performance tests. For example, in a study of Turkish learners of English (Kirkgöz, 2011), the means of pre-tests and post-tests were compared using a *t*-test.

Quantitative analysis was used to study the variety of question types and question strategies used (Abuseileek & Qatawneh, 2013); opinions regarding mobile phone use (Gromik, 2012); “fluency, pronunciation, vocabulary, accuracy and task accomplishment” (Kirkgöz, 2011, p. 4); fluency (13 different measurements) (Kormos & Dénes, 2004); fluency, pronunciation,

complexity, and accuracy (Sun, 2012); and pronunciation and grammar (Tognozzi & Truong, 2009).

Expert evaluation refers to either a researcher or instructor's assessment of the learners' performance. For example, Kirkgöz (2011, p. 4) created a rating scale to assess learner performance in terms of "fluency, pronunciation, vocabulary, accuracy and task completion," which she later used for quantitative comparison. Similarly, in Kormos and Dénes' (2004) study, three native and non-native speakers rated the learners' performance in the AMOC task.

Conclusions on Methodologies

It is puzzling that a majority of studies in this review focused on learner perceptions of language gains without considering expert evaluations or empirical measurements of learner performance. That is, although survey, journal, and reflection data constituted only a marginally larger category than the use of recordings as data, if it were combined with interview data to create the broader category of learner perceptions, it would contain twice as many instances of data collection (n=25) as the recordings category (n=12). It is worth noting that this is a count of instances that each collection method was encountered, where one article may use both surveys and interviews. In other words, researchers relied more heavily on learner perceptions of speech production than on their recorded speech production when studying AMOC in language learning, including studies focusing on the effect of AMOC on learner language gains.

While learner perceptions of linguistic growth and of activity effectiveness are no doubt important aspects in evaluating AMOC and its associated activities, the use of learner perceptions as the sole means of determining this growth and effectiveness is fraught with validity issues. It is doubtful that learners are the best means of gauging language improvement. First, learners are not experts in the language and therefore frequently do not know when they

are saying something correctly or incorrectly. Second, they are not trained in noticing different aspects of their own speech. Finally, they are not trained in reliably rating their linguistic performance.

Learner perceptions may still be of value when combined with other analysis methods. One method is expert evaluation. Native speakers and highly proficient non-native speakers are more familiar with the language and can more accurately determine the quality and accuracy of the learner's performance. Objective measurements, such as words produced per second, will provide even more accurate evidence regarding some aspects of learner performance, such as fluency. Taken together, learner perceptions, expert evaluation, and objective measurements would enable researchers to more accurately evaluate learner language gains from using AMOC.

Conclusions

AMOC can be beneficial to learners in promoting language gains. Studies considered in this review investigated its effects on accuracy, fluency, and pronunciation, showing that it can be a useful technology in helping learners develop these aspects of their language. However, the research does not universally show that AMOC leads to language gains. Additional studies on the effectiveness of using AMOC would enable us to determine with greater reliability whether it is a viable means of promoting language gains. Additionally, the scope of studies should extend beyond grammatical accuracy, fluency, and pronunciation to include such linguistic aspects as complexity, lexical accuracy, and lexical variety (to name a few).

However, we did identify several factors that contribute to effective use of AMOC in a language-learning curriculum. In designing AMOC activities, instructors and designers should consider the learners' access to hardware and software as well as their internet speed. Because many learners are not familiar with recording and editing software, learners will benefit from

technical training. Learners will also benefit from structured self-evaluation and revision cycles, preparatory activities, and project-based learning.

Current research on the effectiveness of AMOC on speaking performance focuses heavily on learner perceptions of language gains. Although learner perceptions can give us clues about their linguistic self-awareness and their experience as AMOC users, they are not an appropriate data source for inferential studies and not the only factor that should be considered by instructors or programs deciding on whether or how to implement AMOC activities. Triangulating with other data sources (such as recordings of learner speech) and other analysis types (such as expert evaluation and empirical measurements) would allow researchers to make more accurate claims as to the effectiveness of AMOC in promoting foreign language gains. This study shows that there are several studies about the qualitative effects of AMOC but few studies providing empirical evidence for linguistic gains through AMOC. What is lacking is an analysis of whether each study's data and analysis type matches the study's claims and conclusions. Such an analysis would help us to better evaluate the trustworthiness of the various conclusions about the usefulness and effectiveness of AMOC.

In this review, audio-based and video-based AMOC were studied together. However, it is not clear if video-based AMOC is more or less effective at promoting language gains when compared to audio-based AMOC. It is possible that video may be detrimental for some learners in that it will likely increase anxiety when compared to audio. On the other hand, video provides a higher fidelity experience when communicating with other learners or the instructor. A purposeful comparison would help determine if the use of either purely audio or purely video-based AMOC is generally most effective, or to which situations and learner types each is best suited.

A final note is that while self-evaluations and revisions promote language gains, it is unclear what systems for self-evaluating and revising are most effective. For instance, is one cycle of video drafting sufficient or must learners follow three or four cycles before they become sufficiently aware and make sufficient revisions? Furthermore, to what degree do learners even follow the specified self-evaluation and review processes? That is, we do not know the extent to which learners revise their recordings after self-evaluating.

AMOC remains an intriguing means of promoting spoken language gains but further research is needed to determine what aspects of spoken language it is best suited for developing and how to effectively incorporate it into a curriculum. AMOC does not appear to be, as some may think, inferior to face-to-face or other synchronous forms of communication. Continued popularity of asynchronous social media, such as Twitter, Snapchat, and YouTube, suggests that it is important to study and understand the unique outcomes and situations where each method can be most useful.

References

- American Council on the Teaching of Foreign Languages. (2012). *Performance descriptors for language learners*. Retrieved from <http://www.actfl.org/publications/guidelines-and-manuals/actfl-performance-descriptors-language-learners>
- Abuseileek, A. F., & Qataweh, K. (2013). Effects of synchronous and asynchronous computer-mediated communication (CMC) oral conversations on English language learners' discourse functions. *Computers and Education, 62*, 181–190.
doi:10.1016/j.compedu.2012.10.013
- Bakar, N. A., Latiff, H., & Hamat, A. (2013). Enhancing ESL learners speaking skills through asynchronous online discussion forum. *Asian Social Science, 9*(9), 224–234.
doi:10.5539/ass.v9n9p224
- Baker-Smemoe, W., Dewey, D. P., Bown, J., & Martinsen, R. A. (2014). Does measuring L2 utterance fluency equal measuring overall L2 proficiency? Evidence from five languages. *Foreign Language Annals, 47*(4), 707–728. Retrieved from <http://doi.org/10.1111/flan.12110>
- Castañeda, M., & Rodríguez-González, E. (2011). L2 speaking self-ability perceptions through multiple video speech drafts. *Hispania, 94*(3), 483–501.
- Clark, R. (1994). Media will never influence learning. *Educational Technology Research and Development, 42*(2), 21–29. doi: 10.1152/advan.00094.2010
- Clifford, R. (2002). *Achievement, performance, and proficiency testing*. Paper presented at the Berkeley Language Center Colloquium on the Oral Proficiency Interview, University of California at Berkley.

- Crookes, G. (1989). Planning and interlanguage variation. *Studies in Second Language Acquisition, 11*(4), 367–383.
- Delaney, T. (2012). Quality and quantity of oral participation and English proficiency gains. *Language Teaching Research, 16*(4), 467–482. Retrieved from <http://doi.org/10.1177/1362168812455586>
- Dixon, E. M., & Hondo, J. (2014). Re-purposing an OER for the online language course: A case study of Deutsch Interaktiv by the Deutsche Welle. *Computer Assisted Language Learning, 27*(2), 109–121. doi:10.1080/09588221.2013.818559
- Dona, E., Stover, S., & Broughton, N. (2014). Modern languages and distance education: Thirteen days in the cloud. *Turkish Online Journal of Distance Education, 15*(3), 155–170.
- Engin, M. (2014). Extending the flipped classroom model: Developing second language writing skills through student-created digital videos. *Journal of the Scholarship of Teaching and Learning, 14*(5), 12–26. doi:10.14434/josotlv14i5.12829
- Fukushima, T. (2002). Promotional video production in a foreign language course. *Foreign Language Annals, 35*(3), 349–355.
- Gleason, J., & Suvorov, R. (2012). Learner perceptions of asynchronous oral computer-mediated communication: Proficiency and second language selves. *Canadian Journal of Applied Linguistics, 15*(1), 100–121.
- Goulah, J. (2007). Village voices, global visions: Digital video as a transformative foreign language learning tool. *Foreign Language Annals, 40*(1), 62–78. doi:10.1111/j.1944-9720.2007.tb02854.x

- Gromik, N. A. (2012). Computers & education cell phone video recording feature as a language learning tool: A case study. *Computers & Education*, 58(1), 223–230.
doi:10.1016/j.compedu.2011.06.013
- Graham, C. (2006). Blended learning systems: Definition, current trends, and future directions. In Bonk, C. & Graham, C. (eds.), *Handbook of blended learning: Global perspectives, local designs* (pp. 3–21). San Francisco, CA: Pfeiffer. doi: 10.2307/4022859
- Hastie, P., Brock, S., Mowling, C., & Eiler, K. (2012). Third grade students' self-assessment of basketball dribbling tasks. *Journal of Physical Education and Sport*, 12(4), 427–430.
Retrieved from <http://doi.org/10.7752/jpes.2012.04063>
- Hirotsani, M. (2009). Synchronous versus asynchronous CMC and transfer to Japanese oral performance. *Calico Journal*, 26(2), 413–438. Retrieved from
<http://doi.org/10.1016/j.cpen.2012.02.001>
- Hirotsani, M., & Lyddon, P. A. (2013). The development of L2 Japanese self-introductions in an asynchronous computer-mediated language exchange. *Foreign Language Annals*, 46(3), 469–490. doi:10.1111/flan.12044
- Hung, S. T. (2011). Pedagogical applications of Vlogs: An investigation into ESP learners' perceptions. *British Journal of Educational Technology*, 42(5), 736–746.
doi:10.1111/j.1467-8535.2010.01086.x
- Jamshidi, R., LaMasters, T., Eisenberg, D., Duh, Q. Y. & Curet, M. (2009). Video self-assessment augments development of videoscopic suturing skill. *Journal of the American College of Surgeons*, 209(5), 622–625. Retrieved from
<http://doi.org/10.1016/j.jamcollsurg.2009.07.024>

- Karweit, N. (1984). Time on task reconsidered: Synthesis of research on time and learning. *Educational Leadership*, 41(8), 32–35.
- Kirkgöz, Y. (2011). A blended learning study on implementing video recorded speaking tasks in task-based classroom instruction. *Turkish Online Journal of Educational Technology*, 10(4), 1–13.
- Kitade, K. (2000). L2 learners' discourse and SLA theories in CMC: Collaborative interaction in internet chat. *Computer Assisted Language Learning*, 13(2), 143–166. Retrieved from [http://doi.org/10.1076/0958-8221\(200004\)13](http://doi.org/10.1076/0958-8221(200004)13)
- Kormos, J., & Dénes, M. (2004). Exploring measures and perceptions of fluency in the speech of second language learners. *System*, 32(2), 145–164. doi:10.1016/j.system.2004.01.001
- Lamy, M.-N., & Goodfellow, R. (1999). “Reflective conversation” in the virtual classroom. *Language Learning & Technology*, 2(2), 43–61.
- Lepore, C. E. (2014). Influencing students' pronunciation and willingness to communicate through interpersonal audio discussions. *Dimension*, 73–96.
- Lin, H. (2015). Computer-mediated communication (CMC) in L2 oral proficiency development: A meta-analysis. *ReCALL*, 27(3), 261–287. Retrieved from <http://doi.org/10.1017/S095834401400041X>
- McIntosh, S., Braul, B. & Chao, T. (2003). A case study in asynchronous voice conferencing for language instruction. *Educational Media International*, 40(1), 63–73.
doi:10.1080/0952398032000092125
- Ono Y., Onishi A., Ishihara M., & Yamashiro M. (2015). Voice-based computer mediated communication for individual practice to increase speaking proficiency: Construction and pilot study. In Zaphiris P. & Ioannou A. (eds), *Learning and collaboration*

- technologies. LCT 2015. Lecture Notes in Computer Science, 9192.* New York, NY: Springer.
- Pop, A., Tomuletiu, E. A., & David, D. (2011). EFL speaking communication with asynchronous voice tools for adult students. *Procedia - Social and Behavioral Sciences, 15*, 1199–1203. doi:10.1016/j.sbspro.2011.03.262
- Sauro, S. & Smith, B. (2010). Investigating L2 performance in text chat. *Applied Linguistics, 31*(4), 554–577.
- Segalowitz, N. (2010). *Cognitive bases of second language fluency.* New York, NY: Routledge.
- Shih, R. (2010). Blended learning using video-based blogs: Public speaking for English as a second language students. *Australasian Journal of Educational Technology, 26*(6), 883–897.
- Sun, Y. C. (2012). Examining the effectiveness of extensive speaking practice via voice blogs in a foreign language learning context. *CALICO Journal, 29*(3), 494–506.
- Sun, Y. C., & Yang, F. Y. (2015). I help, therefore, I learn: Service learning on Web 2.0 in an EFL speaking class. *Computer Assisted Language Learning, 28*(3), 202–219. doi:10.1080/09588221.2013.818555
- Tiraboschi, T., & Iovino, D. (2009). Learning a foreign language through the media. *Journal of E-Learning and Knowledge Society, 5*(3), 133–137.
- Tognozzi, E., & Truong, H. (2009). Proficiency and assessment using WIMBA voice technology. *Italica, 86*(1), 1–23.
- Yaneske, E., & Oates, B. (2010). Using voice boards: Pedagogical design, technological implementation, evaluation and reflections. *Australasian Journal of Educational Technology, 26*(8), 233–250. doi:10.3402/rlt.v18i3.10767

Ziegler, N. (2013). *Synchronous computer-mediated communication and interaction: A research synthesis and meta-analysis* (Unpublished doctoral dissertation). Georgetown University, Washington, DC.

Appendix A

Articles Reviewed in this Study

- Abuseileek, A. F., & Qatawneh, K. (2013). Effects of synchronous and asynchronous computer-mediated communication (CMC) oral conversations on English language learners' discourse functions. *Computers and Education*, *62*, 181–190. Retrieved from <http://doi.org/10.1016/j.compedu.2012.10.013>
- Bakar, N. A., Latiff, H., & Hamat, A. (2013). Enhancing ESL learners speaking skills through asynchronous online discussion forum. *Asian Social Science*, *9*(9), 224–234. Retrieved from <http://doi.org/10.5539/ass.v9n9p224>
- Castañeda, M., & Rodríguez-González, E. (2011). L2 speaking self-ability perceptions through multiple video speech drafts. *Hispania*, *94*(3), 483–501.
- Dixon, E. M., & Hondo, J. (2014). Re-purposing an OER for the online language course: A case study of Deutsch Interaktiv by the Deutsche Welle. *Computer Assisted Language Learning*, *27*(2), 109–121. Retrieved from <http://doi.org/10.1080/09588221.2013.818559>
- Dona, E., Stover, S., & Broughton, N. (2014). Modern languages and distance education: Thirteen days in the cloud. *Turkish Online Journal of Distance Education*, *15*(3), 155–170.
- Engin, M. (2014). Extending the flipped classroom model: Developing second language writing skills through student-created digital videos. *Journal of the Scholarship of Teaching and Learning*, *14*(5), 12–26. Retrieved from <http://doi.org/10.14434/josotlv14i5.12829>
- Fukushima, T. (2002). Promotional video production in a foreign language course. *Foreign Language Annals*, *35*(3), 349–355. Retrieved from <https://doi.org/10.1111/j.1944-9720.2002.tb01860.x>

- Gleason, J., & Suvorov, R. (2012). Learner perceptions of asynchronous oral computer-mediated communication: Proficiency and second language selves. *Canadian Journal of Applied Linguistics, 15*(1), 100–121.
- Goulah, J. (2007). Village voices, global visions: Digital video as a transformative foreign language learning tool. *Foreign Language Annals, 40*(1), 62–78. Retrieved from <http://doi.org/10.1111/j.1944-9720.2007.tb02854.x>
- Gromik, N. A. (2012). Computers & education cell phone video recording feature as a language learning tool: A case study. *Computers & Education, 58*(1), 223–230. Retrieved from <http://doi.org/10.1016/j.compedu.2011.06.013>
- Hirotnani, M., & Lyddon, P. A. (2013). The development of L2 Japanese self-introductions in an asynchronous computer-mediated language exchange. *Foreign Language Annals, 46*(3), 469–490. Retrieved from <http://doi.org/10.1111/flan.12044>
- Hung, S.-T. (2011). Pedagogical applications of Vlogs: An investigation into ESP learners' perceptions. *British Journal of Educational Technology, 42*(5), 736–746. Retrieved from <http://doi.org/10.1111/j.1467-8535.2010.01086.x>
- Kirkgöz, Y. (2011). A blended learning study on implementing video recorded speaking tasks in task-based classroom instruction. *Turkish Online Journal of Educational Technology, 10*(4), 1–13.
- Kormos, J., & Dénes, M. (2004). Exploring measures and perceptions of fluency in the speech of second language learners. *System, 32*(2), 145–164. Retrieved from <http://doi.org/10.1016/j.system.2004.01.001>
- Lepore, C. E. (2014). Influencing students' pronunciation and willingness to communicate through interpersonal audio discussions. *Dimension, 73*–96.

- McIntosh, S., Braul, B., & Chao, T. (2003). A case study in asynchronous voice conferencing for language instruction. *Educational Media International*, 40(1), 63–73. Retrieved from <http://doi.org/10.1080/0952398032000092125>
- Pop, A., Tomuletiu, E. A., & David, D. (2011). EFL speaking communication with asynchronous voice tools for adult students. *Procedia - Social and Behavioral Sciences*, 15, 1199–1203. Retrieved from <http://doi.org/10.1016/j.sbspro.2011.03.262>
- Shih, R. (2010). Blended learning using video-based blogs: Public speaking for English as a second language students. *Australasian Journal of Educational Technology*, 26(6), 883–897.
- Sun, Y.-C. (2012). Examining the effectiveness of extensive speaking practice via voice blogs in a foreign language learning context. *CALICO Journal*, 29(3), 494–506.
- Sun, Y.-C., & Yang, F.-Y. (2015). I help, therefore, I learn: Service learning on Web 2.0 in an EFL speaking class. *Computer Assisted Language Learning*, 28(3), 202–219. <http://doi.org/10.1080/09588221.2013.818555>
- Tognozzi, E., & Truong, H. (2009). Proficiency and assessment using WIMBA voice technology. *Italica*, 86(1), 1–23.
- Yaneske, E., & Oates, B. (2010). Using voice boards: Pedagogical design, technological implementation, evaluation and reflections. *Australasian Journal of Educational Technology*, 26(8), 233–250. Retrieved from <http://doi.org/10.3402/rlt.v18i3.10767>

Appendix B

Comparison of Articles Reviewed

Author(s)	Data Type				Data Analysis Type					
	Surveys, journals, & reflections	Recordings	Interview transcripts	Observation notes	Qualitative analysis	Descriptive measurements	Quantitative comparison	Expert evaluation	Correlation	Unknown/unstated
Abuseileek, A. F., & Qatawneh, K.	✓	✓	✓		✓	✓	✓			
Bakar, N. A., Latiff, H., & Hamat, A.		✓	✓		✓	✓				
Castañeda, M., & Rodriguez-González, E.	✓				✓	✓				
Dixon, E. M., & Hondo, J.	✓				✓	✓				
Dona, E., Stover, S., & Broughton, N.	✓				✓	✓				
Engin, M.	✓		✓		✓					
Fukushima, T.		✓								✓
Gleason, J., & Suvorov, R.	✓		✓		✓	✓				
Goulah, J.	✓	✓		✓	✓					
Gromik, N. A.	✓	✓	✓		✓	✓	✓		✓	
Hirutani, M., & Lyddon, P. A.		✓			✓	✓				
Hung, S.-T.	✓	✓	✓		✓	✓				
Kirkgöz, Y.	✓	✓	✓		✓	✓	✓	✓		
Kormos, J., & Dénes, M.		✓				✓	✓	✓	✓	
Lepore, C. E.	✓				✓	✓		✓	✓	
McIntosh, S., Braul, B., & Chao, T.	✓			✓	✓	✓				
Pop, A., Tomuletiu, E. A., & David, D.			✓		✓	✓				
Shih, R.	✓	✓	✓		✓	✓		✓		
Sun, Y.-C.	✓	✓			✓	✓	✓	✓		
Sun, Y.-C., & Yang, F.-Y.	✓	✓	✓		✓	✓				
Tognozzi, E., & Truong, H.	✓				✓		✓	✓		
Yaneske, E., & Oates, B.	✓		✓		✓					

DELIBERATE PRACTICE OF L2 FLUENCY WITH
SELF AND EXPERT FEEDBACK

Eric H. Young

Brigham Young University

Deliberate Practice of L2 Fluency with Self and Expert Feedback

Abstract

Oral asynchronous computer-mediated communication provides language learners with opportunities to increase their linguistic self-awareness, revise their L2 production, and engage in deliberate practice. In this study, I examined the effectiveness of learner video draft creation and two feedback types on L2 fluency. I used multiple quantitative analysis methods to analyze mean syllable duration, mean length of utterance (in seconds and syllables) and average silent pause duration, each corrected for individual L1 fluency scores. Although in the majority of statistical analyses there were no significant differences, the few that were found suggest that video recording, feedback, and follow-up draft recording may be of some value in helping learners improve their fluency. Additionally, it was found that structural equational modelling may be a more valuable means of analyzing some forms of language learning in the classroom than others due to the challenges of conducting classroom-based learning.

Keywords: asynchronous computer-mediated communication, fluency, foreign language, video, speaking

Introduction

Helping students become more aware of their actual speech production is critical to their ability to adjust their L2 output, increase their willingness to engage in more speaking practice (Lockley, 2013), and become more fluent. But, how many instructors can find the time to meet with each student individually to provide sufficient, personalized, high quality feedback on a regular basis? Burdened by constraints on their time, it is challenging, if not impossible at times, for instructors to give their students that level of feedback. Furthermore, once students leave the safety of their language classes, they will need to rely on themselves for much of their corrective feedback.

Asynchronous computer-mediated communication (ACMC) is a method that can be used to provide learners with L2 speaking practice outside the classroom setting and that can also be used to provide feedback to learners to help them increase their L2 self-awareness.

Asynchronous text, audio, or video communication allows learners to prepare and post their communications at any time and from any place that is convenient to them. Furthermore, because communication does not happen in “real-time,” both the instructor and the learner can review communications before the learner sends it to a classmate or an L2-speaking pen pal, or before submitting a final draft to the instructor. In other words, ACMC provides the instructor with an opportunity to train learners in how to gain linguistic self-awareness on their own and then act on it.

The concern we are left with, though, is how self-awareness raising activities compare to the value of instructor feedback and to what degree learners will benefit from them. That is, instructors have more experience with the L2 and thus a greater sensitivity to differences between learner and native production of the L2. Additionally, they know the common learning

pitfalls for the learners in their classes. On the other hand, instructors may not be as intimately aware of the learners' knowledge and abilities as the learners themselves, they may not know what feedback is best suited to each learner (Huang, 2016), and describing the value of instructor feedback assumes at the outset that learners are reading/listening to their instructor's feedback, which is not always the case.

Since learners will eventually have to provide their own corrective feedback as they develop fluency, and since instructor feedback comes at a higher cost, finding effective ways to train learners to self-evaluate is a pedagogical necessity. The purpose of this study, then, is to use oral ACMC to compare the effectiveness of expert feedback and self-awareness raising activities on L2 fluency.

Literature Review

In this section I will discuss the concept of fluency and its measurements; aspects of deliberate practice; and the role of asynchronous computer-mediated communication in facilitating deliberate practice.

Fluency

There is little disputation that fluency is an important aspect of linguistic ability. Job applications related to foreign languages ask for a fluency rating. Grading rubrics in foreign language classes typically have a category devoted to fluency. Even the mere mention of speaking a foreign language in a conversation almost invariably leads to a question of fluency in that language. But despite how frequently we talk about fluency, it is used to describe a variety of concepts.

Segalowitz (2010) described fluency as “a multidimensional construct” (p. 5) and categorized different measures of fluency into utterance fluency, cognitive fluency, and

perceived fluency. Utterance fluency itself refers to a variety of fluency aspects, perhaps the most famous of which is speech rate (Segalowitz, 2016, p. 81). The main idea behind utterance fluency is a description of produced speech. Perceived fluency also deals with produced speech, but from the standpoint of a listener. In other words, a native speaker's subjective assessment of the learner's fluency (2016, p.86). Cognitive fluency refers to a learner's mental processes while producing and preparing to produce an utterance (2016, p. 82). It is most commonly measured by receptive language tasks.

In her chapter on fluency assessment, De Jong (2016) argued that only two measures of speed fluency (a type of utterance fluency) are not confounded by other measures. In most cases, measures are confounded by pausing, whether or not pausing was the fluency aspect to be measured. Articulation rate, on the other hand, is a factor of the number of syllables and phonation time, where phonation time is total time minus silent pauses. The other measure, mean syllable duration, is the inverse measure of articulation time (i.e., phonation time divided by number of syllables). Mean syllable duration, she argued, provides a more normal distribution than articulation rate.

Deliberate Practice

If fluency is an important aspect of linguistic ability, then activities specifically designed to increase learners' fluency should have a central role in language courses. In numerous studies of different disciplines, Ericsson has referred to these types of activities as deliberate practice (see Ericsson, 2009, for example). In one study on learning and skill acquisition in the medical field, Ericsson (2004) defined deliberate practice as follows:

1. First, the participants were instructed to improve some aspect of performance for a well-defined task.

2. Second, they were able to get detailed, immediate feedback on their performance.
3. Finally, they had ample opportunities to improve their performance gradually by performing the same or similar tasks repeatedly (p. S72).

Highlighting the importance of feedback in deliberate practice in other contexts, Searston and Tangen (2017) showed that university students learning to identify fingerprints performed the best with training and immediate corrective feedback on their performance when compared to learners who received only training. As it relates to language learning, Saito and Akiyama (2017) found that the learners involved in a study of 30 Japanese learners of English, who communicated with native speakers and received immediate feedback in the form of recasts, improved their speech rate (a type of speed fluency) over a 12-week university course.

However, feedback in deliberate practice does not necessarily have to come in the form of external feedback. Ericsson's (2004) described deliberate practice as a solitary activity in which learners spend as much as four hours practicing alone. Although external feedback might provide an otherwise unknown perspective on the learners' performances, the solitary nature of deliberate practice may, in fact, push learners to self-assess and take more responsibility for their learning. According to Ericsson, the efforts of learners in his study "involve[d] problem solving and finding better methods to perform the task" (2004, p. S73).

Asynchronous Videos in Deliberate Fluency Practice

Asynchronous video and audio communication lends itself well to deliberate practice. Learners engaged in producing asynchronous videos or audio recordings are likely to work alone and rehearse their performance multiple times before posting it on social media or submitting it to an instructor, addressing point three from Ericsson (2004), above. In a study of nine university-level Spanish language learners, Castañeda and Rodríguez-González (2011) trained

learners to produce a video responding to a prompt, evaluate their own speech in the video, and then revise their video communication drafts. Learners found that the process helped them to improve their linguistic self-awareness, as well as their complexity, accuracy, and fluency.

The learners in Castañeda and Rodríguez-González's (2011) study were not left to their own self-evaluative devices, though. Rather, the researchers specifically trained the learners to look for grammar, pausing/hesitancy, and pronunciation problems in their oral performance. This training is important because of the difference in skill between experts and non-experts in evaluating performance. As they described this gap, "another challenge related to self-evaluation is learners' difficulty in objectively judging their own performance, compounded by learners' perceived paucity of sufficient knowledge and experience to make judgments about the language they produce" (p. 485).

In a study of 15 university level dance students conducted by Leijen, Lam, Wildschut, Simons, and Admiraal (2009), learners were similarly given guidance in how to reflect on their dance performance. The majority of learners felt that watching and analyzing their videos led to a more realistic view of their dancing. One learner indicated that the responsibility of evaluating his own performance validated previous expert feedback he had received. "Ah, I have heard this so many times before, and you can start to respect what you have been told" (p. 174). But while the expert feedback may be valuable, the responsibility of knowing one's own abilities may be more important and lead to more change. Citing more general feelings from the learners, Leijen et al. stated, "the benefit most often reported was that self-evaluation was more meaningful than merely receiving corrections from the teacher" (p. 174). It is possible, then, that language learners may benefit more from a production, feedback, and revision cycle if the feedback comes in the form of well-trained, learner self-evaluation.

However, most recent studies cited in a review by Young and West (2018) investigating the development of speaking skills through cycles of feedback-informed audio and video production have restricted their data sets and analysis methods to the learners' perceptions of their own linguistic improvement. Without knowing how accurately learners can evaluate their own performance, the only claim such studies can reasonably make are that learners feel they have improved. Although learner experiences are indeed important for other purposes, such as design recommendations, only empirical measures (i.e., recordings of the learners' speech performances) and analysis methods can validate empirical claims.

It is my purpose in this study to compare the effects of guided self-evaluation with expert evaluation on learner speaking performance when creating prompt-based asynchronous videos. I will answer the following research questions:

1. By how much does learners' fluency change by recording a second presentation draft that is informed by linguistic awareness-raising activities?
2. Does self-evaluation or expert evaluation promote more fluency gains in learner-based iterative video production?
3. Does order of treatment matter when learners participate in self-evaluation and are given expert evaluation?

Methodology

In this section I will describe the learners involved in this study, the language task and intervention they will participate in, the fluency measurements I will collect, and my methods for analyzing these measurements.

Participants

Thirty learners in a first semester university Arabic course participated in this study. Learners received 14 weeks of instruction, consisting of 50 minute daily classes and an average of 2 hours per night of homework. Learners were surveyed at the beginning of the semester to determine their linguistic background, whether any are heritage speakers, have participated in StarTalk language programs for high school students, or are retaking this Arabic course. They were also asked about their gender, and other language experience. This study was approved by the IRB. I only used data from participants who gave their consent through an IRB-approved consent form, shown in Appendix A.

Task

Learners created a series of videos, in each of which they responded to a different prompt in Arabic. Instructors were asked to assign each prompt every two weeks starting on the third week, with a planned extra week between videos 5 and 6 due to a holiday break. However, in practice, assignment dates for videos 1 and 3 were pushed back by the instructors, and a technological problem pushed video 2's assignment back as well. The planned schedule and the schedule that the instructors ended up following are shown in Figure 1. Each prompt related to a different topic, coinciding with the textbook they use.

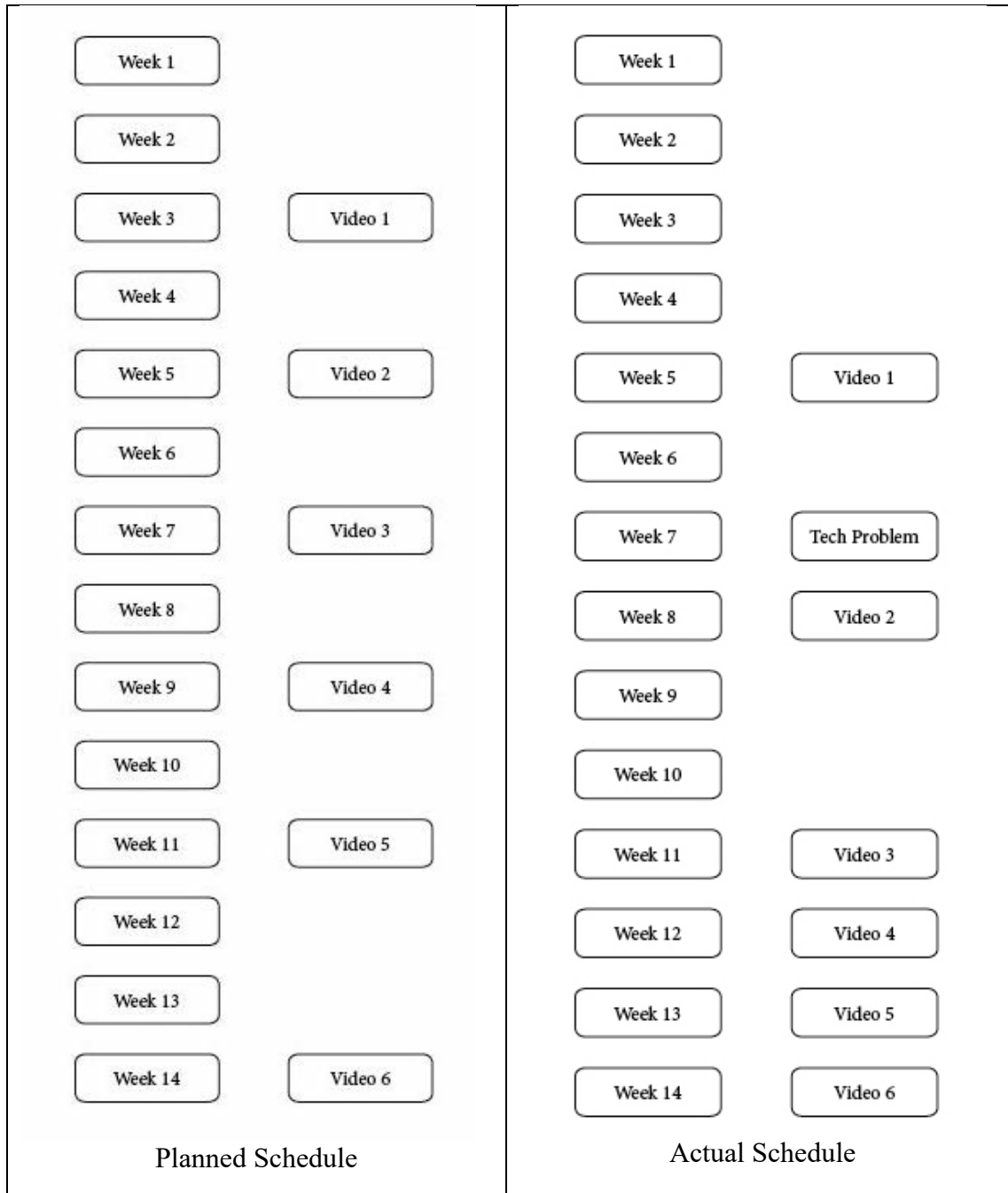


Figure 1. Timing of intervention activities. Video assignments are assigned and submitted on the week they fall next to.

Each video assignment consisted of five major steps.

1. Learners are given a topic-based prompt to consider.

2. Learners watch at least two model videos on the topic assigned, selecting vocabulary expressions for use in their own video.
3. Learners create a video script responding to the prompt and include three vocabulary expressions from the models, then record it.
4. Learners participate in one of two evaluation intervention types (expert or self) according to their group assignment.
5. Learners revise and re-record their video based on their evaluation.

Grouping. Learners were randomly assigned to groups by the process of optimal multivariate matching prior to randomized group assignment. In this process, learners were first paired with a learner of a similar background, based on demographic data obtained through an IRB-approved questionnaire (Appendix B). One learner from each pair was randomly assigned to one treatment group, with the other learner assigned to the second treatment group. There were five exceptions in this process. Two learners from one course section could not be matched with a similar learner and were therefore paired together before being assigned a treatment group. The other course section consisted of an odd number of students, leaving one learner unmatched. This learner was randomly assigned without a partner. Finally, two learners were not available to consent to participate until after the matching took place. These two learners were paired and randomly assigned a treatment group.

In a case study involving 132 participants and 14 covariates, Greevy, Lu, Silber, and Rosenbaum (2004) validated this process by running 10000 matched and 10000 unmatched trials. They found that matched randomizations were, statistically speaking, equivalent to having a 7% larger sample size, and that using matched randomizations avoided some pitfalls

experienced with unmatched randomizations (2004). In our current study, learners were paired according to the following information:

- gender (i.e., male, female)
- age (as a continuous variable, e.g., 20 years old)
- year in school (i.e., 1 for freshman, etc.)
- major of study (included as three variables: Middle Eastern Studies, linguistics, other)
- course section (i.e., section 1, 2, 3, or 4)
- previous Arabic experience
- previous experience with other foreign languages

Learners in each group participated in all five of the steps outlined in the Task subsection above. However, group A conducted self-evaluation for the first three videos and then switched to expert evaluation, while group B did the opposite. Table 1 illustrates these assignments.

Table 1

Learner Grouping and Evaluation Types

Group	Video 1	Video 2	Video 3	Video 4	Video 5	Video 6
Draft	1 2	1 2	1 2	1 2	1 2	1 2
A	Self	Self	Self	Expert	Expert	Expert
B	Expert	Expert	Expert	Self	Self	Self

Expert evaluation. A teaching assistant fluent in Arabic provided the 30 learners with text-based feedback on their videos. Feedback was based on the pronunciation of learners' selected expressions, the frequency of learner pausing while speaking, and the frequency of repairs they make while speaking (i.e., saying an incorrect word and immediately following up with the correct or intended word). Feedback came in the form of the assistant annotating the

learners' videos, indicating the location of pauses, repairs, and instances of unsatisfactory pronunciation, and describing the problem and providing guidance on how to improve.

Self-evaluation. Learners were trained to analyze their own videos, comparing the pronunciation of their selected expressions with the pronunciation of those expressions in the model videos, to notice and explain the reason for their pauses, and to notice and explain the reason for their repairs.

Measurement

As this study is an empirical analysis of learner speech, the learners' video files themselves acted as the unit of analysis. To prepare these files, I first converted them into audio format. Then, silences at the beginning and end of each recording were trimmed so as not to confound measurements. Finally, I used Praat, a phonetic analysis tool, to measure two aspects of utterance fluency: speed fluency and breakdown fluency.

I measured speed fluency in terms of mean syllable duration (MSD), which is calculated as phonation time / n syllables. De Jong (2016) identified 11 commonly used measures of speed fluency, arguing that all but two are confounded with other measures. The two exceptions were articulation rate and its inverse, mean syllable duration. De Jong further argued for the superiority of mean syllable duration over articulation rate due to the fact that articulation rate produces a more skewed distribution.

Unlike utterance fluency, De Jong (2016) did not advocate for a clear best measurement of breakdown fluency. To account for this, I selected both mean length of utterance (MLU) and average silent pause duration (i.e., total time – phonation time / number of silent pauses), or ASPD, since (a) they are not confounded by other measures; (b) they can be measured automatically rather than by hand; and (c) having two measures instead of one may provide a

more accurate description of learner breakdown fluency. Additionally, I used two different measures of the mean length of utterance: time, in seconds (MLUsec) and syllables (MLU_{syll}). This is because learners may be able to increase the number of syllables they produce between pauses even if they do not increase the temporal length of their speaking.

Based on De Jong and Bosker's (2013) validation of a pause measurement instrument, I set the lower pause duration threshold at 300ms. By using the upper end of the threshold range De Jong and Bosker presented, I hoped to account for the fact that participants in this study are in their first semester of university level Arabic and may have slightly longer pauses.

Additionally, Baker-Smemoe, Dewey, Bown, and Martinsen's (2014) contended that the use of 300ms as a lower threshold avoids limitations found in previous fluency studies involving pause duration. Their study measured the fluency of L2 speakers of five languages, ranging from novice to superior level proficiency according to the ACTFL proficiency scale (American Council on the Teaching of Foreign Languages, 2012).

In addition to collecting these measurements for each draft, I collected baseline L1 fluency measurements. De Jong, Groenhout, Schooen, and Hulstijn (2015) showed the importance of accounting for learners' individual speaking style when assessing L2 fluency. To do so, she proposed correcting L2 fluency measures based on L1 fluency measures. Thus, prior to participating in either treatment group, learners in my study recorded a video draft in which they responded to a prompt in their native language. I collected the three measurements of their speed fluency and breakdown fluency described above as baseline L1 fluency data to compare against their developing L2 fluency.

Analysis

To answer question 1, *by how much does learner fluency change by recording a second presentation draft that is informed by linguistic awareness-raising activities?* I conducted a paired samples *t* test, comparing the change in fluency from the first draft to the second draft for each of the six video assignments for each learner. This was done separately for each measurement of fluency (i.e., MSD, MLUsec, MLUsyll, and ASPD), normalized according to De Jong, et al. (2015) procedure: L1 fluency measurement minus L2 fluency measurement, e.g., $L1\ MSD - L2\ MSD = nMSD$ (p. 238). A *p* value of $<.05$ indicates a significant change in learner fluency when creating a second draft in conjunction with linguistic awareness-raising feedback. An effect size was also calculated to determine the degree to which learners increased their fluency for all significant results.

To answer question 2a, *does self-evaluation or expert evaluation promote more fluency gains in learner-based iterative video production?* I first conducted an independent samples *t* test, comparing the change in each of the normalized fluency measures between self-evaluation and expert feedback treatment types. I used the fluency measurements taken from learners' first draft of the first video and second draft of the third video assignment, which is the last draft learners recorded that is not confounded by exposure to multiple treatment types.

Following the independent samples *t* test, I conducted a linear regression on each of the corrected fluency measures using the seven pairing variables listed in the Grouping subsection above as covariates. Additionally, we used completion rate (i.e., amount of treatment) as a covariate. Although the *t* test provides a stronger indication of whether one treatment is superior, a linear regression provides a better model of the relationships between learner fluency and the stated covariates. Thus, I explored whether either treatment is best for a given type of learner.

It was hypothesized that expert evaluation benefits learners with more experience (i.e., older, higher year in school, more previous language experience) the most. Experienced learners are more likely to understand the goal of corrective feedback and apply the feedback to more aspects of their spoken performance. Less experienced learners, on the other hand, may make the specific changes indicated by the TA, but they may not apply the feedback to other aspects of their performance. However, it is not expected that the self-evaluation treatment benefits either the more or less experienced learners differently, because learners in either group may have higher aptitude or experience with self-evaluation. Additionally, it is expected that learners who complete more video assignments will benefit more in terms of fluency gains.

To answer question 2b, *are either effective in the first place?* A latent growth model (LGM) was created. LGM was selected because this study involves repeated measures (i.e., two drafts per video assignment), where participants with different linguistic backgrounds switch treatment types halfway through the intervention. Additionally, whereas ANOVA and regression analyze mean change and “treat differences among individual subjects as error” (Duncan & Duncan, 2009, p. 979), LGM accounts for both group and individual growth. This model allowed me to compare the initial level of learning with the amount of learning taking place at the end of each treatment.

To answer the third question, *does order of treatment matter when learners participate in self-evaluation and are given expert evaluation?* I analyzed the growth of learners’ fluency measures over time using a latent growth model (LGM). LGM was selected because it accounts for differences in individual background (e.g., number of languages known), it tracks both individual learning paths and group learning, and because it is able to work with incomplete data sets. The LGM shows the change in the amount and rate of learning that takes places, both as

affected by learners doing more videos and learners switching treatment types. Similar growth curves for both treatment types would indicate that they are effectively the same. However, a different rate of change for learning at the point of switching treatment types would suggest that one treatment type is more effective than the other at building fluency.

Results

In this study, I aimed to find out: (a) how much learners increased their fluency by recording a second presentation draft informed by linguistic awareness-raising activities; (b) whether self-evaluation or expert evaluation promoted more fluency gains in learner-based iterative video production; and (c) whether order of treatment mattered when learners participated in self-evaluation and were given expert evaluation. To answer these questions, I measured the MSD, MLUsec, MLUsyll, and ASPD of six cycles of two student-recorded video drafts (total of 12 videos). Each measurement was normalized (nMSD, nMLUsec, nMLUsyll, and nAPSD, respectively) based on the participants' L1 fluency, as described above. In this section, I discuss the results of the statistical tests conducted to answer these questions. Results have been condensed, where possible, to make them clearer for the reader. However, non-significant results have been reported to help keep the significant results in perspective.

Fluency Change from First to Second Drafts

To determine whether the fluency measurements were normally distributed, skewness was divided by the standard error of skewness. Any resulting quotients above 2 or below -2 were not considered normal distributions. Fluency measurements were normally distributed in 37.5% of the videos. The quotients are displayed in Table 2. Each assignment was first analyzed with a parametric test. Results of non-parametric tests were compared with the

parametric test results to confirm the findings. In all cases, both the parametric and non-parametric tests agreed on the significance of results.

Table 2

Quotients of Skewness and Standard Error of Skewness as an Assessment of Distribution

Video	Draft 1	MSD	MLUsec	MLU _{syll}	ASPD
1	1	1.911	5.985	5.449	1.990
	2	1.408	2.208	2.729	-0.536
2	1	2.512	2.323	1.858	0.873
	2	0.243	1.002	1.569	1.590
3	1	1.467	1.453	2.693	1.127
	2	1.656	1.210	1.124	0.819
4	1	4.567	2.355	2.823	0.983
	2	0.767	5.102	4.611	2.073
5	1	0.453	5.323	3.069	1.827
	2	0.066	1.372	1.804	2.076
6	1	1.258	1.441	2.603	1.190
	2	-0.377	4.653	5.274	0.385

A paired-samples *t* test was calculated to compare the fluency of first and second drafts of each video in terms of nMSD. No significant difference was found ($p > .05$), implying there was no measureable improvement across the drafts. A Wilcoxon paired-samples test for non-parametric data confirmed these results. The mean, standard deviation, and *p* values of the paired-samples *t* test for each video are listed in Table 3.

Table 3

Results of Paired-Samples t Tests for nMSD

Video	<i>p</i> Value	df	Mean	Std Dev
1	.190	20	.006	.021
2	.122	12	.008	.017
3	.800	13	-.002	.024
4	.880	14	.001	.032
5	.847	13	.002	.029
6	.218	12	.005	.015

A paired-samples *t* test was calculated to compare the fluency of first and second drafts of each video in terms of nMLUsec. A significant difference was found for video 4 ($p = .046$, $d = .566$). The effect size, d , was calculated as mean difference/standard deviation, where .5 represents an increase of half a standard deviation. This is a moderate effect size according to Cohen (1988, p.26). Growth, or mean change in normalized utterance length, was calculated as nMLUsec draft 2 (-2.6664) - nMLUsec draft 1 (-2.9309), or .265 seconds. Where utterance length is longer in native speech, on average, learners improved by .265 seconds after completing video 4. No other significant difference was found. A Wilcoxon paired-samples test for non-parametric data confirmed these results. The mean change, standard deviation, and p values of the paired-samples *t* test for each video are listed in Table 4.

Table 4

Results of Paired-Samples t Tests for nMLUsec

Paired-Samples <i>t</i> Test				
Video	<i>p</i> Value	df	Mean Difference	Std Dev
1	.585	20	-.082	.673
2	.293	12	-.117	.384
3	.444	13	.069	.325
4	.046	14	.265	.467
5	.643	13	.057	.447
6	.337	12	.195	.702

A paired-samples *t* test was calculated to compare the fluency of first and second drafts of each video in terms of nMLUsyll. A significant difference was found for video 4 ($p = .042$, $d = .577$). The effect size, d , was calculated as mean difference/standard deviation, where .5 represents an increase of half a standard deviation. This is a medium effect size according to Cohen (1988, p. 26). Growth, or mean change in normalized utterance length, was calculated as nMLUsyll draft 2 (-11.356) - nMLUsyll draft 1 (-12.528), or 1.173 syllables. Where utterance length is longer in native speech, on average, learners improved by 1.173 syllables after completing video 4. No other significant difference was found. A Wilcoxon paired-samples test for non-parametric data confirmed these results. *P* values for each video are listed in Table 5. The mean change, standard deviation, and *p* values of the paired-samples *t* test for each video are listed in Table 5.

Table 5

Results of Paired-Samples t Tests for nMLUsyll

Video	Paired-Samples <i>t</i> Test			
	<i>p</i> Value	df	Mean Difference	Std Dev
1	.603	20	-.310	2.690
2	.516	12	-.274	1.474
3	.554	13	.220	1.353
4	.042	14	1.173	2.032
5	.280	13	.480	1.591
6	.305	12	1.226	4.124

A paired-samples *t* test was calculated to compare the fluency of first and second drafts of each video in terms of normalized ASPD. A significant difference was found for video 3 ($p = .019$, $d = .714$). The effect size, d , was calculated as mean difference/standard deviation, where .7 represents an increase of over half a standard deviation. This is a medium effect size according to Cohen (1988, p. 26). Growth, or mean change in normalized silent pause duration, was calculated as nASPD draft 2 (.101) - nASPD draft 1 (.232), or -.131 seconds. Where silent pause duration is shorter in native speech, on average, learners improved by .131 seconds after completing video 3. No other significant difference was found. A Wilcoxon paired-samples test for non-parametric data confirmed these results. The mean change, standard deviation, and p values of the paired-samples *t* test for each video are listed in Table 6.

Table 6

Results of Paired-Samples t Tests for nASPD

Video	Paired-Samples t Test			
	p Value	df	Mean Difference	Std Dev
1	.182	20	-.070	.231
2	.798	12	-.012	.165
3	.019	13	-.131	.183
4	.071	14	-.129	.255
5	.306	13	-.058	.205
6	.069	12	-.096	.174

The Effect of Self- and Expert Evaluation on Fluency

An independent samples t test was calculated to compare the mean change in fluency of participants who completed first and final video drafts of each treatment type in terms of nMSD, nMLUsec, nMLUsyll, and nASPD. To do this, the mean change was calculated (i.e., mean fluency at video 3 draft 2 minus mean fluency at video 1 draft 1) for each group and then compared between groups. No significant difference was found. A Mann-Whitney U test for non-parametric data confirmed these results. The mean, standard deviation, and p values of the independent samples t test for each phase of treatment are listed in Table 7. Levene's test showed that the variances for each treatment were equal for each fluency measure.

Table 7

Results of Independent Samples t Tests for the Change in Fluency for the First Treatment Period

Measure	N	Levene's Test		Independent Samples t Test	
		p Value	p Value	Mean Diff.	Std Dev
MSD	15	.355	.186	-.019	.026
MLUsec	15	.139	.619	-.149	.541
MLU syll	15	.178	.866	-.202	2.156
ASPD	15	.889	.727	.053	.274

Simple linear regressions were calculated to predict the change in the learners' normalized fluency at the end of the first treatment period (i.e., from video one draft one to video three draft two) based on how many of the six total video drafts they completed. The ANOVA F-test for the regression equations were not significant. Completion rate was not a significant predictor of change in normalized fluency for the first treatment period. Regression equations are shown in Table 8.

Table 8

Regression Equations for Change in Normalized Fluency Measures from First to Last Video of First Treatment Period Covaried by Completion Rate

Measure	Equation	p Value	R ²
nMSD	F(1,13) = .362	.558	.027
nMLUsec	F(1,13) = .002	.963	.000
nMLU syll	F(1,13) = .012	.916	.001
nASPD	F(1,13) = .421	.528	-.043

Multiple linear regressions were calculated to predict the change in the learners' normalized fluency at the end of the first treatment period (i.e., from video one draft one to video three draft two) based on age, year in school, major of study, treatment group, gender, previous Arabic experience, and previous non-Arabic foreign language experience. The ANOVA F-test for these regression equations for normalized fluency were not significant. The factors tested in this study were not significant predictors of change in normalized fluency for the first treatment period. The results for the ANOVA F-test for the regression equations are shown in Table 9.

Table 9

ANOVA F-test for the Regression Equations for Change in Normalized Fluency Measures from First to Last Video of First Treatment Period Covaried by Age, Year in School, Major of Study, Treatment Group, Class, Gender, Previous Arabic Experience, and Previous Non-Arabic Language Experience

Measure	Equation	<i>p</i> Value	R ²
nMSD	F(8,6) = .366	.905	.328
nMLUsec	F(8,6) = .472	.839	.386
nMLU _{syll}	F(8,6) = .317	.932	.297
nASPD	F(8,6) = .372	.902	-.560

An LGM was constructed to compare the effectiveness of the treatments in terms of learner fluency at the end of the first treatment period (i.e. before the treatments were switched). *P* values for the slope of each fluency measure are displayed in Table 10. A significant slope was found for nMSD ($p = .020$). Normalized MSD for learners in the self-feedback group changed at a rate of -.009 per video draft, whereas nMSD for learners in the TA feedback group changed at a rate of -.004 per video draft. The effect size of this slope (i.e., 6 total video drafts in this part of the treatment process * change in nMSD) is -.054 seconds for learners in the self-

feedback group, and -.024 seconds for learners in the TA feedback group. Since fluent syllables take less time to produce, movement in the negative direction indicates improvement. Where average syllable duration across all learners for video 1 draft 1 is .245 seconds, -.054 seconds represents a 22% improvement and -.024 represents a 10% improvement, with learners in the self-feedback group improving at a faster rate.

Table 10

Results of LGMs Exploring the Effect of Treatment Type on Fluency for Videos 1–3

Measure	<i>p</i> Value
nMSD	.020
nMLUsec	.549
nMLUsyll	.306
nASPD	.708

To further explore factors predicting fluency, an LGM was constructed to analyze the effect of completion rate of all 12 video drafts on fluency growth. In other words, I regressed completion on the slope and intercept coefficient of the LGM. *P* values for the slope of each fluency measure are displayed in Table 11. A significant effect of completion on slope was found for nASPD ($p = .049$). For every 1 unit increase in completion rate the slope decreased by .023 ($p = .049$), compared to an increase of .015 for learners who did not complete the video. Thus a person who completed all videos would see a -.096 second ($(-.023 + 0.015) * 12$) decrease (improvement) in their fluency across the entire course, compared to a .18 second ($0.015 * 12$) increase in silent pause duration for learners who completed no videos. In practical terms, -.096 seconds represents a 10% improvement, whereas .18 seconds represents a 18.7% worsening, where the average ASPD across all learners for video 1 draft 1 is .962 seconds.

Two distinct periods can be seen in the plot of the nASPD model shown in Figure 2. The first period includes time points 0, 1, 2, and 3, representing video 1 draft 1 through video 2 draft 2. This period does not contain a clear pattern. This may be attributed to spacing of the assignments. As shown in Figure 1, above, roughly two and a half weeks separated video 1 from video 2, as well as video 2 from video 3, due to implementation and technology challenges. By contrast, learners were able to access the remaining videos on a weekly basis.

The second period includes the remaining time points, from video 3 draft 1 through video 6 draft 2. Here, two clear patterns emerge. The first is the general increase in nASPD that we found in the LGM. The second is a consistent decrease in nASPD from draft 1 to draft 2 of each video. This period of the plot suggests both an improvement (from draft to draft) and a worsening in nASPD (overall) based on the amount of videos each learner completed. It should also be noted that some participants who were interviewed about their video recording experience indicated going about their preparations and recordings in a different manner as they were starting out compared to later videos. Due to the finding of a significant result, the noticeable difference between the two periods on the plot, and due to the qualitative evidence that at least some learners changed their preparation and recording style around this time, I decided to further explore this finding by creating models based only on drafts from videos 3 through 6.

Table 11

Effects of Completion Rate on the Slopes of the LGMs on Their Respective Outcomes

Outcome Measure	<i>Effect of Completion Rate on Slope</i>	<i>p Value</i>
nMSD	0.003	.205
nMLUsec	-.001	.985
nMLUsyll	-.141	.523
nASPD	-.023	.049

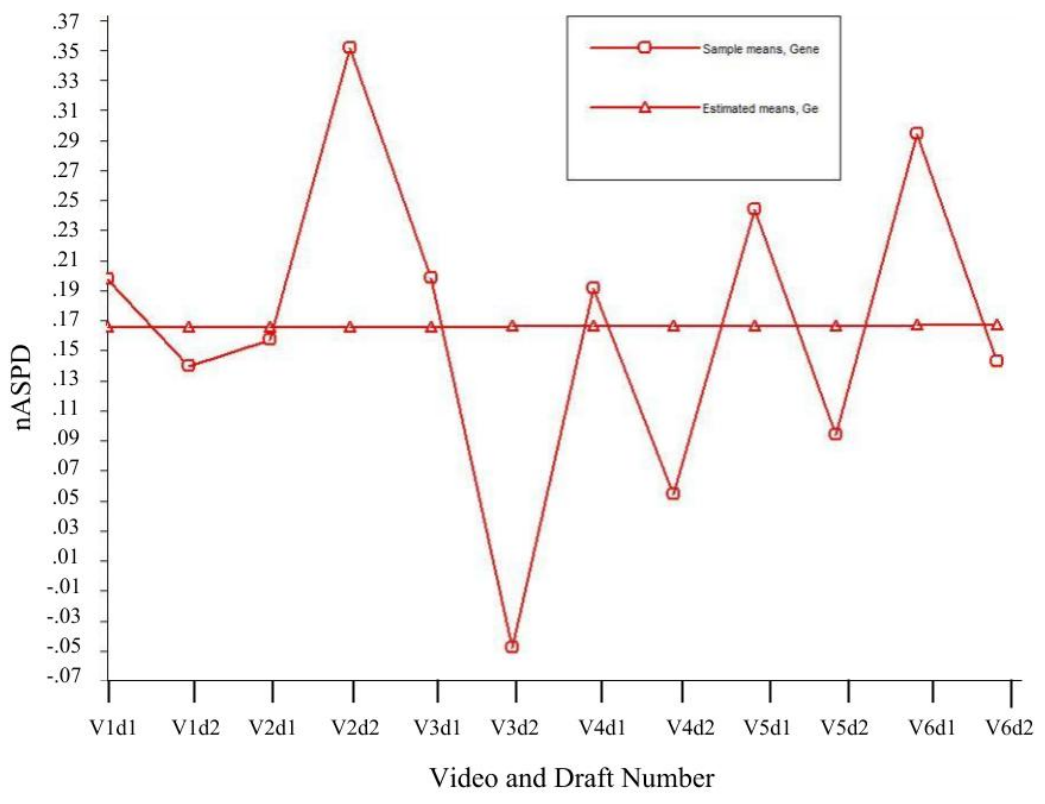


Figure 2. Plot for the LGM of normalized ASPD based on completion rate.

An LGM was constructed to explore the effect of completion rate on fluency for videos 3 through 6. P values for the slope of each fluency measure are displayed in Table 12. No significant result was found for any of the four measurements.

Table 12

Results of an LGM Exploring the Effect of Completion Rate on Fluency for Videos 3–6

Measure	P Value
nMSD	.420
nMLUsec	.696
nMLUsyll	.634
nASPD	.157

To provide further information about the effect of video recording assignments on fluency irrespective of other factors, I constructed an LGM tracking the change in learner fluency from videos 1 through 6 and again from videos 3 through 6. P values for the slope of each fluency measure are shown in Table 13. A significant slope was found for nMLUsec ($p = .044$) from videos 3 through 6. The effect size was calculated at an increase of .037 seconds of nMLU per video, or .296 seconds after all eight videos in this model. The result for nMLUsyll, although not statistically significant at the 95% level, was nonetheless interesting ($p = .062$).

Table 13

Results of LGMs Tracking Unconditional Fluency-Change

Measure	<i>Videos 1 – 6</i>	<i>Videos 3 – 6</i>
	<i>p Value</i>	<i>p Value</i>
nMSD	.837	.711
nMLUsec	.532	.044
nMLUsyll	.877	.062
nASPD	.279	.224

Effectiveness of Order of Treatment

An LGM was constructed to analyze the effectiveness starting with either treatment type, comparing the fluency of each group at the end of both treatment periods (i.e., video 6 draft 2). Learners in group A engaged in self-feedback for videos 1 – 3, then switched to TA feedback for videos 4 – 6. Learners in group B engaged in TA feedback first, then self-feedback. *P* values for the effect of the order of treatment on the slope of the LGM are displayed in Table 14. No significant result was found for any of the four measurements.

Table 14

Results of an LGM Comparing the Effect of the Order of Treatment from Videos 1–6 on the Slope of the LGM

Measure	<i>p Value</i>
MSD	.066
nMLUsec	.229
nMLUsyll	.285
nASPD	.076

Based on the reasoning in the previous section and the patterns found in Figure 2, I followed up on this question by constructing an LGM to compare the change in fluency for learners in each treatment group for videos 3 through 6. *P* values for the slope of each fluency measure are displayed in Table 15. A significant slope was found for nASPD ($p = .017$). Normalized ASPD for learners in group A changed at a rate of $-.037$ seconds per video draft, whereas nASPD for learners in group B changed at a rate of $-.016$ (or $-.037 + .021$) seconds per video draft. The effect size of this slope (i.e., 8 total video drafts * change in nASPD) is $-.296$ seconds for learners in group A, and $-.124$ seconds for learners in group B. Where average silent pause duration across all learners for video 1 draft 1 is $.962$ seconds, $-.296$ seconds represents a 31% improvement and $-.124$ seconds represents a 13% improvement, with learners engaged in self-feedback first decreasing (i.e., improved) their silent pause duration more rapidly than learners who received TA feedback first.

Table 15

Results of an LGM Comparing the Effects of Order of Treatment for Videos 3–6 on the Slope

Measure	<i>P</i> value
nMSD	.9
nMLUsec	.165
nMLUsyll	.757
nASPD	.017

Discussion

In this section I discuss the findings from the study in relation to each of the three research questions.

Fluency Change from First to Second Drafts

Although there were significant results from the paired-samples *t* tests and Wilcoxon paired-samples tests, these occurred in only one video each for nASPD, nMLUsec, and nMLUsyll, with a medium effect size for each. This amounts to a total of three significant results out of 24 total statistical tests. On one hand, the paucity of significant results suggests that very little change is taking place in learner fluency as a direct result of the video recording assignments. On the other hand, the fact that there are significant results, particularly with a medium effect size, despite the small population size and short treatment period, suggests that perhaps growth would be seen under different conditions.

However, whereas the Wilcoxon and paired-samples *t* tests could only provide results for learners who completed each pair of recordings, the follow-up LGM for nASPD referred to in the third section of the results and in Figure 2 indicated that learners did, in fact, benefit from recording video drafts. The results of this analysis show noticeable and consistent improvement (31% or 13% for group A and B, respectively) from each first draft to its respective second draft for video assignments three to six. This level of change can be perceived by listeners and not only through computer measurement. This provides quantitative support for Castañeda and Rodríguez-González' (2011, p. 491) qualitative findings that draft recordings help learners improve their fluency.

The general trend of worsening nASPD across all drafts (not from first to second drafts) can at least be partially attributed to the fact that learners were asked to remember more and more vocabulary and use increasingly more complex structures as the course progressed. Thus, learners would need to balance cognitive resources between remembering their script, new

vocabulary and structures, and older/partially-learned vocabulary and structures, pausing at greater length to remember what to say next.

The Effect of Self- and Expert Evaluations on Fluency

Similar to what took place in the first research question, neither the non-parametric nor the parametric tests revealed significant differences between treatment types, which may suggest they are equivalent, or at least that we are unable to prove a difference at this point. This could be attributed to the small sample size, the short treatment time, or a genuine lack of difference between treatment benefits. However, the LGM revealed an improvement in learner syllable duration for both treatment groups. Unlike the *t* tests, the LGMs were able to work with data from learners who did not complete the first or final draft of the treatment, predicting growth based on fluency change trajectories. Although learners who received either type of feedback reduced their syllable duration, those who provided their own feedback made greater improvements (22% compared to 10%).

That self-feedback is comparably better than receiving expert feedback is a partial surprise. One reason for this surprise is that learners may be less sensitive to their own disfluencies in a foreign language. One of Ericsson's (2004) conditions for deliberate practice is that learners receive *immediate* feedback (p. S72). Although expert feedback may be more detailed and sensitive, the immediate nature of the self-feedback process may have allowed them to incorporate it more easily into their second drafts, rendering it more valuable.

Additionally, of the covariates studied, only completion rate had a significant effect on fluency. The lack of effect by demographic data is not entirely surprising, as it did not include specific details about learners' language learning background, such as the amount of time spent studying other languages or their language aptitude. However, the finding that completion rate

affected fluency is in line with the general notion that time on task is a critical factor in language learning (Clifford, 2002). In this case, the finding was limited to one aspect of fluency, ASPD, and was modest at a 10% improvement for those completing all video assignments. It is possible that the small number of video drafts relative to the number of other class assignments learners were responsible for masked the effect of completion rate on the other fluency measures. Thus, it may be worth conducting a follow-up study to focus on the effect of completion rate on fluency improvement.

Effectiveness of Order of Treatment

Regarding question 3, no differences in fluency were found based on the order that learners engaged in either type of feedback in the initial analysis. However, after focusing on the final four video assignments (eight videos in all, including each draft) in the LGM, it was revealed that learners who engaged in self-feedback activities before moving on to receiving expert feedback improved at a greater rate than those who received expert feedback first. This suggests that learners need to gain experience in the feedback process in order to understand the value of expert feedback and how to better implement it.

Limitations

Many issues limit the generalizability of these research findings. One of which was the sample size. Although I attempted to account for this through a priori matching and the use of latent growth modelling, it is difficult to conclude that a similar study would produce the same findings.

A second limitation was the focus on fluency based on rehearsed speech. This study does not account for differences between rehearsed and spontaneous speech. Any improvements that learners made in terms of their rehearsed speech may not apply to their spontaneous

conversations. In fact, fluency by itself may be too narrow a measure, as tradeoffs leading to growth in one dimension of language acquisition occur at the expense of other dimensions (Skehan, 2009). Since learners in this study felt compelled to produce more accurate speech with each iteration of their production, it would be valuable to determine if both accuracy and fluency are improving together in this particular situation or if any fluency growth has led to a loss of accuracy or complexity.

Yet another limitation was the nature of the feedback. Learners were provided with text-based feedback for pronunciation, pausing, and repairs. Although a simple text-based note explaining the presence of a pause might suffice, text-based descriptions of pronunciation likely led to less understanding of the feedback and less improvement of pronunciation.

Additionally, more than 90% of the learners had previously studied a non-Arabic foreign language, and many of them lived in a country speaking that language for between one and a half and two years. This background is uncommon, even for university students. Replication of this study may, therefore, lead to different results.

Perhaps more limiting are the realities of implementing the video recording activities themselves. Instructors provided slightly different instructions and levels of support for the assignments, which led to different levels of participation and understanding of the instructions. Additionally, the level of script-writing varied between videos. Learners seldom wrote a script for their English videos, typically wrote verbatim scripts for the earlier Arabic videos, and typically wrote looser scripts using bullet points and key words for their later videos, as revealed in follow-up interviews.

Although the focus of this study is learner fluency gains, it should be noted that a limitation of this study is the fact that learners are required to mimic native speaker models and

intentionally include multi-word vocabulary items. A pure focus on fluency without vocabulary constraints would likely lead to greater fluency. However, inclusion of vocabulary was a requirement of the classes involved in this study.

To overcome some of these limitations, it is recommended that future research include the following modifications:

- Increase the sample size by including multiple class levels.
- Measure the level of correlation between change of fluency in rehearsed speech due to feedback types with change in fluency of spontaneous speech.
- Increase the sample size by studying the same course across several semesters.
- Increase the usability of feedback by providing it in recorded format, perhaps including written feedback with it when deemed necessary.
- Increase the length of study by spanning multiple semesters.
- Increase the uniformity of experience by providing better training for instructors.

Although instructors involved in this study were given training, it was not done in a formal environment and they were not given any paperwork to refer back to.

Formalizing the training and providing written resources might lead to better implementation of the intervention.

Implications

There are three major implications of this study. First, there appears to be some value in engaging in recording video presentations, recording follow-up drafts, and engaging in feedback and linguistic awareness-raising activities. Instructional designers should consider the inclusion of asynchronous video activities as part of their foreign language curriculum because of the

opportunities they afford learners to practice outside the classroom at their convenience, and because of the activities' generally positive effect on learner fluency.

Second, the fact that there were few significant results despite the small sample size and short length of treatment suggests that more research is not only needed but that an improved study could yield more clear results regarding the practical value of these activities. Further studies should increase the number of video recording assignments so that the effect of other classroom assignments is mitigated. Further studies should also track the change in learner fluency across a longer period of time, such as across multiple semesters.

The third major implication is the value of latent growth modeling in research involving classroom-based foreign language learners. The use of LGM in language learning research is not unprecedented. It has been used to predict academic achievement based on L1 proficiency (Guglielmi, 2008), to track the growth of ESL learner paragraph writing skills (Aryadoust, 2016), and to study the development of morphological and vocabulary knowledge for Spanish-speaking English language learners (Kieffer & Lesaux, 2012).

Latent growth modeling and its ability to deal with incomplete data sets may be more useful in classroom-based foreign language studies because of the challenges in implementing research methods. Students regularly neglect some assignments, deciding whether to prioritize other assignments, work, or social activities. Additionally, many foreign language classes are small, consisting of 8 – 15 students. This is especially true of less commonly taught languages. In some cases, having each student participate in the study on classroom activities can be crucial to statistical power. The value of the LGM is that it can make up for some of these challenges by including more of the available data.

References

- American Council on the Teaching of Foreign Languages. (2012). *Performance descriptors for language learners*. Retrieved from <http://www.actfl.org/publications/guidelines-and-manuals/actfl-performance-descriptors-language-learners>
- Aryadoust, V. (2016). Understanding the growth of ESL paragraph writing skills and its relationships with linguistic features. *Educational Psychology, 36*(10), 1742–1770. Retrieved from <http://doi.org/10.1080/01443410.2014.950946>
- Baker-Smemoe, W., Dewey, D. P., Bown, J., & Martinsen, R. A. (2014). Does measuring L2 utterance fluency equal measuring overall L2 proficiency? Evidence from five languages. *Foreign Language Annals, 47*(4), 707–728. Retrieved from <http://doi.org/10.1111/flan.12110>
- Castañeda, M., & Rodríguez-González, E. (2011). L2 speaking self-ability perceptions through multiple video speech drafts. *Hispania, 94*(3), 483–501.
- Clifford, R. (2002). *Achievement, performance, and proficiency testing*. Paper presented at the Berkeley Language Center Colloquium on the Oral Proficiency Interview, University of California at Berkley.
- Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Hillsdale, NJ: Lawrence Erlbaum.
- De Jong, N. H. (2016). Fluency in second language assessment. In D. Tsagari & J. Banerjee (Eds.), *Handbook of second language assessment* (pp. 203–218). Boston, MA: De Gruyter Mouton.
- De Jong, N. H., & Bosker, H. R. (2013). Choosing a threshold for silent pauses to measure second language fluency. In R. Eklund (Ed.), *Proceedings of the 6th Workshop on*

- Disfluency in Spontaneous Speech (DiSS)* (pp. 17–20). Retrieved from http://www.isca-speech.org/archive/diss_2013/dis6_017.html
- De Jong, N. H., Groenhout, R., Schoonen, R., & Hulstijn, J. H. (2015). Second language fluency: Speaking style or proficiency? Correcting measures of second language fluency for first language behavior. *Applied Psycholinguistics*, *36*(2), 223–243.
<http://doi.org/doi:10.1017/S0142716413000210>
- Duncan, T. E., & Duncan, S. C. (2009). The ABC's of LGM: An introductory guide to latent variable growth curve modeling. *Social and Personality Psychology Compass*, *3*(6), 979–991. Retrieved from <http://doi.org/10.1111/j.1751-9004.2009.00224.x>
- Ericsson, K. A. (2004). Deliberate practice and the acquisition and maintenance of expert performance in medicine and related domains. *Academic Medicine*, *79*(10), S70–S81. Retrieved from <http://doi.org/10.1097/00001888-200410001-00022>
- Ericsson, K. A. (2009). Enhancing the development of professional performance: Implications from the study of deliberate practice. In K. A. Ericsson (Ed.), *Development of professional expertise* (pp. 405–431). New York, NY: Cambridge University Press.
- Greevy, R., Lu, B., Silber, J. H., & Rosenbaum, P. (2004). Optimal multivariate matching before randomization. *Biostatistics*, *5*(2), 263–275. Retrieved from <http://doi.org/10.1093/biostatistics/5.2.263>
- Guglielmi, R. S. (2008). Native language proficiency, English literacy, academic achievement, and occupational attainment in limited-English-proficient students: A latent growth modeling perspective. *Journal of Educational Psychology*, *100*(2), 322–342.

- Huang, S. C. (2016). Understanding learners' self-assessment and self-feedback on their foreign language speaking performance. *Assessment & Evaluation in Higher Education*, 41(6), 803–820. Retrieved from <http://doi.org/10.1080/02602938.2015.1042426>
- Kieffer, M. J., & Lesaux, N. K. (2012). Development of morphological awareness and vocabulary knowledge in Spanish-speaking language minority learners: A parallel process latent growth curve model. *Applied Psycholinguistics*, 33(1), 23–54. Retrieved from <http://doi.org/10.1017/S0142716411000099>
- Leijen, A., Lam, I., Wildschut, L., Simons, P. R. J., & Admiraal, W. (2009). Streaming video to enhance students' reflection in dance education. *Computers and Education*, 52(1), 169–176. Retrieved from <http://doi.org/10.1016/j.compedu.2008.07.010>
- Lockley, T. (2013). Exploring self-perceived communication competence in foreign language learning. *Studies in Second Language Learning and Teaching*, 3(2), 187–212. Retrieved from <http://doi.org/10.14746/ssllt.2013.3.2.3>
- Saito, K., & Akiyama, Y. (2017). Video-based interaction, negotiation for comprehensibility, and second language speech learning: A longitudinal study. *Language Learning*, 67(1), 43–74. Retrieved from <http://doi.org/10.1111/lang.12184>
- Searston, R. A., & Tangen, J. M. (2017). Training perceptual experts: Feedback, labels, and contrasts. *Canadian Journal of Experimental Psychology*, 71(1), 32–39. Retrieved from <http://doi.org/10.1037/cep0000124>
- Segalowitz, N. (2010). *Cognitive bases of second language fluency*. New York, NY: Routledge.
- Segalowitz, N. (2016). Second language fluency and its underlying cognitive and social determinants. *International Review of Applied Linguistics in Language Teaching*, 54(2), 79–95. Retrieved from <http://doi.org/10.1515/iral-2016-9991>

Skehan, P. (2009). Modelling second language performance: Integrating complexity, accuracy, fluency, and lexis. *Applied Linguistics*, 30(4), 510–532. Retrieved from <http://doi:10.1093/applin/amp047>

Young, E. H., & West, R. E. (2018). Speaking practice outside the classroom: A literature review of asynchronous multimedia-based oral communication in language learning. *EuroCALL Review*, 26(1), 59–78. Retrieved from <http://doi.org/10.4995/eurocall.2018.8599>

Appendix A

IRB-Approved Consent Form

Implied Consent

My name is Eric Young, I am a graduate student at Brigham Young University and I am conducting this research under the supervision of Professor Charles Graham, from the Department of Instructional Psychology & Technology. You are being invited to participate in this research study of “Deliberate Practice of L2 Fluency with Self- and Expert Feedback”. I am interested in finding out about the effectiveness of teaching foreign language learners to provide feedback on their own speech production.

Your participation in this study will require the completion of the attached questionnaire. This should take approximately three minutes of your time. Your participation will be confidential. We may contact you in the future for additional information about your experiences participating in this study. You will not be paid for being in this study. This survey involves minimal risk to you. The benefits, however, may impact society by helping increase knowledge about foreign language learning feedback effectiveness.

You do not have to be in this study if you do not want to be. You do not have to answer any question that you do not want to answer for any reason. We will be happy to answer any questions you have about this study. If you have further questions about this project or if you have a research-related problem you may contact me, Eric Young at 801-830-6595 or Eric.Young@byu.edu or my advisor, Dr. Charles Graham at 801-422-4110 or charles.graham@byu.edu.

If you have any questions about your rights as a research participant you may contact the IRB Administrator at A-285 ASB, Brigham Young University, Provo, UT 84602; irb@byu.edu; (801) 422-1461. The IRB is a group of people who review research studies to protect the rights and welfare of research participants.

The completion of this survey implies your consent to participate. If you choose to participate, please complete the attached survey during class. Thank you!

KEY INFLUENCES OF LEARNER-CREATED ASYNCHRONOUS
VIDEOS ON GAINING L2 FLUENCY

Eric H. Young

Brigham Young University

Key Influences of Learner-Created Asynchronous Videos on Gaining L2 Fluency

Abstract

Integrating technology into the classroom can be beneficial but does not guarantee improvements in learning. Incorporating learner feedback into design and redesign can increase the likelihood of success. In this study, I used the Critical Incident Technique to explore the incidents, influences, and aspects of a video-based learning activity and how they contributed to both learner successes and challenges in developing second language fluency. Interview and survey responses highlighted key design considerations and provided specific feedback about learner experiences. Based on this feedback, I provided key considerations for designing asynchronous video recording assignments: (a) determine the focus, content, and format of feedback given to the learners; (b) provide learners with clear linguistic models; (c) provide a variety of training types at multiple points; (d) set up the assignments in ways that encourage practice; and (e) create realistic, engaging prompts that align closely with course content.

Keywords: critical incident technique, asynchronous computer-mediated communication, foreign language

Introduction

Gaining spoken fluency in a new language is a daunting task. To achieve advanced proficiency in a Romance language, an adept learner needs a minimum of 700 hours of instruction (Blake, 2013). Advanced speakers can talk about familiar topics clearly, but struggle with abstract discussions and unfamiliar topics, and show patterns of errors (ACTFL, 2012; ILR, n.d.). For non-Romance languages, it has been estimated (Blake, 2013; Language Testing International, n.d.) that learners require between 1300 and 2400 hours to achieve the same proficiency, depending on the language. In contrast, a 14-week university course may provide as many as 168 hours, between classroom instruction and homework (assuming 4 classroom hours per week and 2 homework hours per classroom hour). However, during classroom time, speaking practice must compete with teacher explanations, grammar drills, reading, writing, listening, and class announcements. Speaking practice is also hard to coordinate outside the class, due to learners' other courses and extracurricular activities.

The internet and current technologies provide learners with the opportunity to increase the amount of time they spend speaking outside the classroom. As Blake (2013) stated, in support of the use of technology in L2 learning, "technology, then, if used wisely, can play a major role in enhancing L2 learners' contact with the target language" (p. 2). Internet services can provide learners with L2 pen pals. Cell phone apps can facilitate both synchronous video chats and asynchronous video messaging. And these technologies not only provide learners with the opportunity to increase their speaking time, but also allow them to engage in it at the time and place of their choosing rather than being limited to the classroom or language lab.

But the mere existence of such technologies does not guarantee effective language learning. In his meta-analysis of 25 articles studying the use of computer-mediated

communication (CMC) to promote speaking proficiency, Lin (2015) found conflicting evidence of CMC's effectiveness. Although his meta-analysis of CMC use showed a generally small, positive effect compared to face-to-face learning, some of the studies included in the meta-analysis showed a negative effect.

The reason for the conflicting findings can, perhaps, be explained by Clark's (1994) argument that, technology and media only act as the vehicles to pedagogical practice, and it is those practices that determine the effectiveness of learning. He reiterated his position, saying, "any necessary teaching method could be designed into a variety of media presentations" (p. 22). Based on Clark's statements, the mere inclusion of technologies that facilitate communication is insufficient. Instead, learners need carefully crafted learning activities that make use of these technologies.

Since technology-based L2 communication affords learners both more opportunities for speaking practice and greater convenience, this leads to the question of what can improve the effectiveness of technology-based L2 speaking activities. In this study, I turn to learner feedback on a series of learner-created video-based activities to explore the aspects, influences and incidents related to the activities students themselves believe are effective or ineffective.

Critical Incidents and User Feedback

Designers can best create learning environments and activities when they understand how learners interact with those environments and activities. Walker (2015), in a study exploring ESL student perspectives of learning, argued for the importance of user feedback in design, comparing the situation to the customer service industry. He claimed that by "identify[ing] attributes that determine service quality within specific service types . . . the information could assist service managers to focus their efforts, set priorities, and thus improve the efficiency and

effectiveness of their service operations” (p. 98). Learner feedback, then, not only helps the designer understand the learner’s experience as an end unto itself, but it can be instrumental in providing more satisfying and effective learning experiences.

However, not all feedback is useful. Key to Walker’s (2015) argument is the word “specific” (p. 98). In a study on evaluations of students in a military flight school described by Flanagan (1954), it was found that much of the feedback provided by evaluators amounted only to vague descriptions and judgments. He highlighted comments, such as “lack of inherent flying ability” and “insufficient progress,” (p. 328) that do not clearly explain the problem or its origins. Clearly this type of feedback is not specific enough to help a designer improve learner experiences.

Reed (2012) encountered a similar feedback problem in his study of distance education design feedback. She described her participants as “satisfied customers” (p. 61) who left generally positive but ambiguous feedback that was not useful in making improvements to the course design. While positive feedback can be useful, in the case of Reed’s respondents, the majority of the feedback did not specifically cite *what* made the course effective for them. One feedback example that the author considered useful specifically cited a need for more course readings: “I was interested in the writing about scaffolding, but it was not enough...” (p.62).

Specificity is not, by itself, a sufficient criterion for feedback usefulness. The various activities, incidents, and aspects of a design cited in the feedback should be critical to the learner’s success or failure. In a study of instructional design professionals’ design practices, Sugar and Luterbach (2016) discovered several practices that were considered critical to creating good designs. In one piece of feedback, a designer highlighted the importance of collaborating with stakeholders: “Perhaps the most effective instructional design activity is working with

faculty or the SME to ensure a well-thought out and planned organizational structure is in place within their course or learning module” (p. 302).

The Critical Incident Technique (CIT) was developed during Flanagan’s studies of flight school evaluations to “determine the job requirements critical for success” (Butterfield, Borgen, Amundson, & Maglio, 2005, p. 477) and to produce specific reasons for judgments and feedback. This determination is based on attributes of the design, influences of the design on the learner, and specific incidents that led to judgments of success or failure (p. 490). However, for CIT to be useful in improving learning experiences, CIT research should explain where these attributes, influences, and incidents originated from. That is, CIT research should “focus on eliciting the beliefs, opinions and suggestions that formed part of the critical incident” (Butterfield et al., 2005, p. 490) in an effort to “[explore] what helps or hinders in a particular experience or activity” (Butterfield, Borgen, Amundson, & Maglio, 2009, p. 268).

Learning Task

In this study, I explored the critical incidents, influences, and aspects (hereafter referred to simply as incidents) of Arabic language learners creating, analyzing, and revising videos for building their fluency. Thirty learners from two first semester Arabic classes (different instructors) participated in six video activities, each consisting of the same seven steps listed below:

1. Receive a topic-based prompt
2. Watch model videos in which native speakers respond to the prompt or speak about the related topic
3. Extract vocabulary expressions from the model videos
4. Create a video script incorporating expressions

5. Record a first draft reciting (but not reading) the script
6. Participate in a linguistic awareness-raising activity, described later
7. Record a second, revised draft

Learners were expected to complete all seven activities for each video they created, six videos in all. In all, learners completed 64% of the video recordings, with most learners completing all the steps. Miscommunications between the researchers and one instructor led to a lower-than-expected participation in that instructor's class. Before participating in this study, learners were expected to create a video in English to provide personalized baseline L1 fluency data. All learners completed this recording.

During the analysis phase, learners were asked to participate in a linguistic awareness-raising activity as part of a program to train them to become more self-sufficient learners and increase the amount of effective time they can practice speaking outside the classroom. They were divided into two groups, with group A conducting guided self-evaluations for the drafts of their first three videos, then receiving expert evaluation feedback from the course TA for the next three video drafts. Group B participated in the activities in reverse order, receiving expert evaluation from a TA for the first three videos, then conducting self-evaluation for the next three.

Learners in both groups either recorded or uploaded their videos to an app called GoReact. Figure 1 shows GoReact's interface. The interface allows learners and TAs to replay the videos and mark them with three different markers: vocabulary (yellow), pauses (red), and repairs (blue). To mark the videos, learners and TAs click on the marker buttons in the center right-hand side of the interface. Markers are displayed on the video timeline on the bottom left as well as on the video window itself on the top left. Learners and TAs can attach comments to each marker by typing them in the field on the top right.

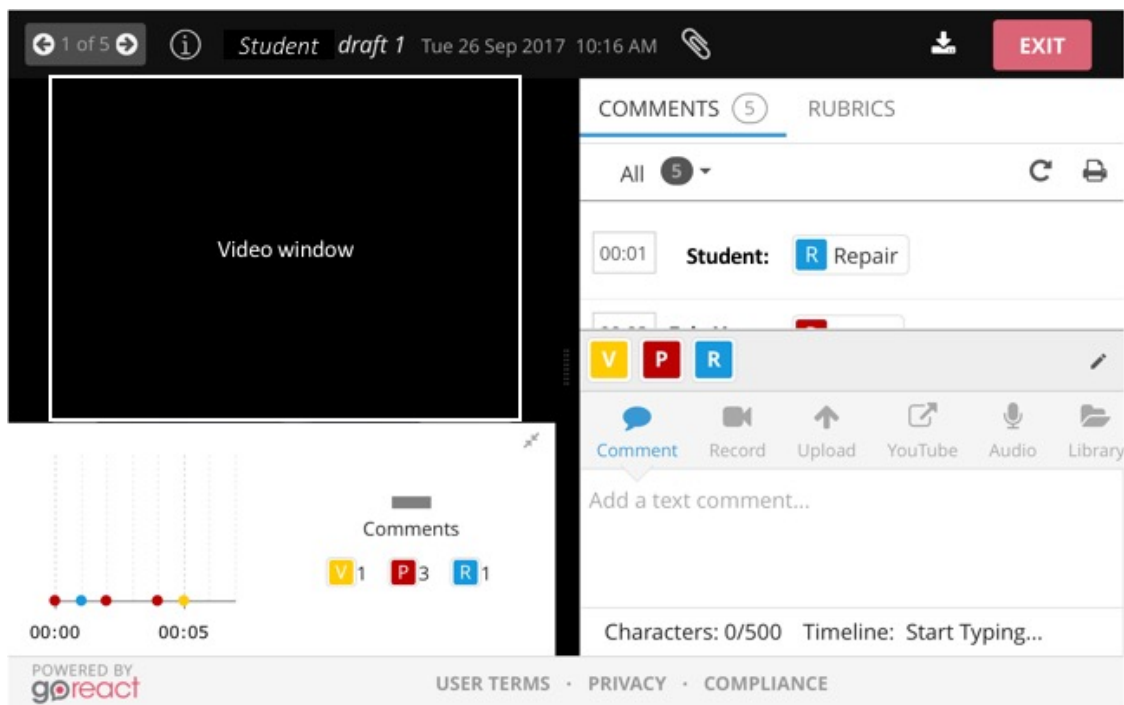


Figure 1. GoReact interface where learners mark (V)ocabulary/pronunciation, (P)auses, and (R)epairs and write comments.

Methods

In this study, I explored the critical incidents of learners' asynchronous video creation activities to find out what about them is effective and ineffective in building their linguistic self-awareness and L2 fluency and in order to make improvements to the activities. Participants were drawn from both sections of a beginning level university Arabic course. All learners in the course (n=30) were expected to participate in the intervention, as it was part of the curriculum. Learners were surveyed at the beginning of the course to find out if they had previous experience learning foreign languages, including Arabic. Few (20%) had prior Arabic experience. Most (90%) had some prior foreign language experience. Twenty five of the thirty learners were in either their first or second year at the university. Four learners were in the Middle Eastern

Studies and Arabic major, three were in the linguistics major, and the remaining twenty-three were in other majors.

After the deadline for the fifth video activity, I interviewed learners to identify the critical incidents of their video activities that they believe rendered the activities effective or ineffective for them. Or, as Flanagan (1954) described the purpose, the “aspects of behavior which are believed to be crucial in formulating a functional description of the activity” (p. 337). I purposively selected a mixture of outlying learners, half who did well and half who did not, expecting that outliers would identify more salient, critical incidents of the activities. As Butterfield, et al., (2005) explained, adequate coverage of learner experiences is determined by a saturation of the number of incidents found rather than by a strict number of interviews (p. 479). I interviewed eight learners from the two sections as well as two learners from a pilot study in which learners participated in a nearly identical instructional task for a total of 10 learners. Most interviews began before the sixth video assignment to ensure that interview time did not compete with learner final exams or post-semester travel.

One of the aims of CIT interview questions is to put learner opinions in context (Butterfield, et al., 2009, p. 270). To this end, I structured interview questions to focus on specifics related to both the activities and the learners’ actions while participating in those activities. The following questions were used in the interviews:

1. Walk me through a typical video assignment.
2. What do you think were the most important things that led to your success on these assignments?
3. What do you think were the most important things that made these assignments challenging?

4. What about that made the assignment more helpful/challenging?
(Note that this question will involve more probing, depending on the interviewees' responses).
5. Were there other things related to the assignments specifically that made the experience more or less effective for you?
6. Was there anything in your own preparation or the ways you personally went about the assignment that made it more or less effective for you?
7. How did you know what you needed to work on?
8. Did you notice a difference between doing your own self-evaluation and getting feedback from the TA?
9. Is there anything about these assignments that you would change or that would make it more effective in helping you become a more fluent speaker?

For each interviewee, I coded and analyzed the interview transcripts, and categorized the incidents I found. Similar to Butterfield, et al.'s (2009) second interview, I sent each interviewee a copy of their transcript via email, along with a list of the themes I identified in their interview. This provided an opportunity to elicit feedback on the incidents and their categorization, and to clarify any issues that arose during analysis (p. 276).

As an additional crosscheck, all learners were then asked to participate in a survey/questionnaire as a follow-up to the interviews. Each learner was given a list of themes common to the different interviews. They were asked to rank-order the themes and to explain their reasoning for their rankings. The following questions were used in the survey/questionnaire:

1. Please rank-order the following incidents, influences, and aspects from most important to least important.
2. Can you explain why you ordered them this way?

Descriptive statistics of the rank-ordering were used to determine which incidents of the video activities are most important to users and should be addressed in future designs or redevelopment. Statements from the surveys provided insights on how to improve the activities.

Critical Incident Technique

CIT, as it is currently practiced, involves nine steps, functioning as credibility checks. The checks, based on Butterfield, et al.'s (2005) summary and Butterfield, et al.'s (2009) description of how to apply the checks, are listed below with an explanation of how they were implemented in this study.

Interview recording. Interviews were recorded to ensure the accuracy of the data. For one interviewee, both recording devices failed halfway through the interview. Many incidents were recovered because I had taken detailed notes, though the details of the interviewee's experience in that half of the interview were lost.

Interview fidelity. At multiple points during this study, a colleague familiar with CIT verified that I followed CIT interview protocol sufficiently.

Exhaustiveness. According to Butterfield, et al.'s (2005), this occurs when "only two or three critical behaviours emerge from 100 critical incidents gathered" (p. 487). In line with this recommendation, I followed a five-step process to ensure that I identified possible themes. First, I identified a total of 588 incidents mentioned in the interviews. Second, I sorted all incidents into 22 major categories. Third, I combined multiple instances of themes into a single entry. For example, 8 instances of "TA feedback" were combined with 8 instances of "getting TA

feedback”. Fourth, I combined themes that were similar in meaning to reduce the list to 48 total themes. For example, “getting feedback on pronunciation”, “getting feedback on fluency/pausing” were combined into “focus of TA feedback”. Fifth, I removed any themes that I, as the designer, had little or no control over and focused on incidents that could be changed in order to improve the video recording assignments. This was based on Flanagan’s (1954) considerations for developing headings and categories. He recommended that “headings used for classification and reporting of the data should be such that findings in terms of them will be easily applied and maximally useful” (p. 345). Thus, although “having a native speaker model” was critical in many participants’ mind, having the native speaker models was one of the parameters set by the Arabic program for these video recording assignments. Likewise, “importance of Arabic relative to other classes” was eliminated from the list because it is not something I could directly influence. Exceptions to this criterion, however, were “your level of understanding the video models (i.e., listening comprehension)” and “your level of Arabic knowledge.” Although I cannot directly influence these influences, they were described as being so important to the learners’ experience that it would be unwise to make any design revisions without considering them.

Participation rates. Flanagan (1954) recommended the reporting of incidents “having significant frequencies” (p. 345). In this study, all major categories (n=20) were mentioned by at least two of the 10 interview participants. That is, in addition to being retained during the sorting process, more than one interviewee considered each of the categories worth mentioning. Although a 20% frequency of mention is not a high rate, the themes mentioned are useful in describing the learners’ experience. For instance, *instructor’s classroom feedback on video recording assignments* highlights an important disconnect between the recordings and the

classroom activities that affected the learners and should be addressed. On average, each category was mentioned by five interviewees. Table 1 shows the percentage of interviewees that mentioned each category.

Table 1

Critical Incident Categories and the Percentage of Interviewees Mentioning Each Category

Critical Incident Category	Frequency of Mention in %
1. Amount of practice learning the assignment before doing it for credit	90%
2. Doing your own error correction using your level of Arabic knowledge (i.e., without TA feedback)	90%
3. Doing error correction based on the TA's feedback	80%
4. Prioritizing task completion (e.g., filling 1 minute of speech) over quality	70%
5. Alignment of video recording assignments and classroom activities	60%
6. Content of TA feedback (including specificity)	60%
7. How well you understood the video models and your ability to extract the vocabulary chunks	60%
8. Timeliness of feedback	60%
9. Alignment of the videos and the prompts	50%
10. Amount of time it took to complete video recording assignments	50%
11. Clarity of assignment instructions	50%
12. Timing of when you did first and second drafts (i.e., space between drafts)	50%
13. Amount of time you worked with other students on these assignments in/out of class	40%
14. Authenticity/quality of the prompts	40%
15. Format of feedback: online vs. in person; text vs. video	40%
16. Importance of video recording assignments in class	40%
17. Quality of demonstrations	40%
18. Transcribing your pronunciation	30%
19. Instructor's classroom feedback on video recording assignments	20%
20. The number of things you were considering during review/revisions	20%

Independent extraction of incidents. I hired and trained a peer to independently extract incidents from three of the ten interviews for comparison with my extracted list. Incidents were

approximately 70% the same. Discrepancies took the form of one coder finding something the other had not already found. These differences led to opportunities to read the interview transcripts from a new perspective and add additional incidents that I had not initially included. There were no disagreements regarding the inclusion of extracted incidents.

Independent coding. I hired and trained an independent coder to categorize the list of extracted incidents from three randomly selected interviews. In contrast to my categories, the independent coder's categories were broad. Because this study was intended to be used to improve learning experience rather than being a broad, exploratory study, we generally opted for the specificity of my categorization style. When considering this difference in approach, however, we were able to make improvements to the list of specific categories based on the list of broad categories. An earlier and longer version of the list containing both specific and broad categories can be found in the Appendix.

After comparing the two categorization styles, we clarified some categories and replaced others entirely. For instance, "online nature of feedback" was broadened to "format of feedback" despite the general push for specific categories. We did this to allow participants to give feedback on more aspects of the feedback process. "Having the TA's feedback as a model for doing your own feedback" was changed to "doing error correction based on the TA's feedback" to focus on the act of error correction. "Consistency of assignment dates" was removed altogether in lieu of "timeliness of feedback." This is because the instructional design called for consistent and regular assignment dates, and assignments were not set consistently due to implementation and technological problems. I believed we could learn more from the learners' beliefs regarding the time it took to give video feedback to them. The resulting list was clearer for the learners and more useful for future design iterations.

Crosschecking. In this study, the second interview (follow-up email) and the survey acted as two crosschecks. In the second interview, interviewees were provided with a transcript coded with the themes identified in their interview so they could certify the accuracy and validity of any conclusions. Three learners responded, indicating that the transcripts and coding were acceptable. For the survey, 13 learners provided responses, three of whom also participated in the interviews.

In a typical survey, this would constitute a low number of respondents. However, according to Butterfield, et al. (2005), the purpose of CIT is to adequately represent the object of study, regardless of the number of participants. This is in line with Flanagan's (1954) focus on the use of expert observers without indicating a need for a certain quantity of expert observers. Although different in methodology in some respects, the Delphi Method, as described by Hasson, Keeney, and McKenna (2000) corroborates the viability of depending on only a small group of participants through the use of experts and multiple rounds of careful investigation. Thus, the present study is based on a sufficient number of survey respondents because it

1. it makes use of multiple layers of investigation;
2. the survey respondents consist of both learners who also participated in the interview, and learners who did not participate in the interview, leading to a broader coverage of perspectives;
3. learners gained some expertise through spending several hours on their assignments.

Expert opinions. I submitted the list of categories to the course instructors and to the Arabic program's section head to judge whether the categories were (a) useful; (b) surprising; (c) lacking any other important categories (Butterfield, et al., 2009, p. 278). One instructor commented that "clarity of assignment instructions" was critical to his students' success. On the

other hand, he pointed out that a practice video assignment during class time would have reduced the learners' stress levels and helped them understand the assignment process better.

Theoretical agreement. Categories found in this study were compared to findings in previous literature and are reported in the discussion section. New categories may suggest either that the category is not sound, or that a new theme has been discovered (Butterfield, et al., 2009, p. 278).

Findings

In this section I discuss the feedback that interviewees and survey respondents provided regarding their video recording assignments.

Interviews

Most of the incidents that interviewees provided feedback on fell into three major categories: (a) nature of the recording assignments; (b) components of the recording assignments; and (c) training. Although technological issues and features were another major category, none of the findings related to this category were surprising or revelatory. Key findings for each major category will be discussed in their respective subsections below.

Nature of the recording assignments. One of the comments regarding the nature of the assignments was a criticism of doing monologs and a desire to instead do dialogs or real conversations. The act of conversation requires participants to listen and respond, rather than focus exclusively on their own linguistic production. In the interviewee's view, neither the Arabic course nor the video recording assignments have sufficiently prepared him to engage in dynamic conversations: he is more comfortable with writing and presentational speaking.

Another interviewee commented on the value of being able to work at her own pace. Learners were typically given the assignment on a Monday morning, expected to complete the

first draft by Wednesday morning, and do revisions and a second draft by Friday morning. The interviewee valued this aspect of the assignments because she could read the prompt, think about a possible response for a day or two, then complete the draft. As a normal college student with multiple classes, the flexibility allowed her to schedule her time in a way that suited her, and provided her with an opportunity to plan.

Yet another aspect of the assignments was the length requirement for each video draft. Learners were expected to memorize and record a full minute of speech for all videos, even at the beginning of the semester. One interviewee reported this as a problem, particularly at the beginning of the semester. She felt she did not know enough Arabic to be able to speak for an entire minute, where much of their course content up to that point revolved around greeting, leave-taking, and listing likes and dislikes. Additionally, remembering that much speech in a new language was difficult. A further complication, reported by another interviewee, was that a minute-long recording on the first draft did not equate to a minute-long second draft. In her experience, there were fewer pauses in the second draft, resulting in a shorter, more condensed recording. A third interviewee described the requirement in positive terms, stating that it prepared her for other speaking tasks. They were “helpful because I was already speaking for a minute long in these videos and so then I was, I guess, a bit more confident when preparing for our speaking assignments,” where speaking assignments consisted of partners meeting with the instructor in person each week.

Components of the recording assignments. One of the first components of the recording assignments that learners see is the prompt. Each assignment has a unique prompt that the learner must respond to. Prompts are written in English to keep them clear.

Interviewees reported that recording prompts do not always align with class activities or the model videos. Indeed, this was a major design challenge because chapters in the course textbook do not follow themes. Rather, they are designed around a single, model video using actors. In line with that, the Arabic program at this university developed additional videos, using the chapter vocabulary and grammar structures, to provide a greater variety of content that was still accessible. In the university's videos, native, non-actor Arabic speakers relate their real stories to the learners. Due to the use of real stories, and because the textbook did not follow clear themes, it was not possible to provide learners with prompts that led learners to create recordings that were equivalent to the models. This, in turn, led to some learners having a hard time creating their responses and finding suitable vocabulary from the model videos.

After reading the prompts and formulating a basic idea of what to say, learners watched up to four model videos before writing a script for their own recording. Learners extracted vocabulary expressions they felt applied to their recording assignment, tailoring them as needed. In the words of one interviewee regarding the value of the model videos, "I could see where the words fit together and what forms of the words they used and how they pronounced them." However, some learners experienced trouble using the video models because they had trouble understanding them. One interviewee described her experience, saying that she was "too busy figuring out what they were saying and not how they were saying it."

After watching the model videos, most learners wrote a script to follow while recording their drafts. Some wrote their scripts verbatim. In the case of one interviewee, this led to him reading his script rather than reciting it from memory, despite the instructions. Other learners opted for rough scripts. In one instance, the interviewee strategically wrote her rough script based on her Arabic abilities. "I'd have like phrases that I wanted to use, things that I knew

might be just more difficult kind of like on the fly, like verbs to conjugate. Things like that.” In another instance, the interviewee wrote his script in English so that the ideas would be more accessible to him, helping him to focus on conveying his meaning: “so that way I can still be like not just reading it in Arabic, but just still translating and talking.” In fact, some interviewees began by writing scripts verbatim at the beginning of the course and then switched to partial scripts or no script at all once they felt confident enough in their language skills to do so. The following is an excerpt from one interview describing his transition through these phases.

But then towards the end I tried to think of what I wanted to say just in English and not translate it and not write it out word for word. And then I'd just try to free speak it and just try to come up with the sentences as I went with the direction but word by word just making it up. And I felt like that helped me. The first drafts of that one were a little bit more rough, and I felt like, more choppy. But I feel like it helped me to make the vocabulary mine a little bit more. And then I felt like by the second draft I was feeling a little bit more comfortable because I wasn't just memorizing the words, but I was using them.

One of the most defining aspects of the recording assignments was the revision process and marking system, and this is reflected in the frequency (160 of 588 incidents) with which it was cited in the interviews. The value of this marking and comment system was that learners had a visual reference of where their error took place, and they could write down feedback specific to any given moment of their recording. “It helped me to see just kind of with the color coding and going through and marking it. I could see the things where I was making mistakes the most in.”

The marking and comment system, assignment structure, and GoReact app, helped learners to increase their linguistic self-awareness by enabling and guiding them to carefully and purposively watching themselves and compare their speech to model Arabic speech. One learner expressed this sentiment in her surprise at watching her video: “Wow. I paused like a lot in this video. I should do it again.” Another described how his increased awareness led him to improvement. “I’d go back over it and mark the places where I’d paused and maybe the words that maybe had the intention of being right but came out sounding a little weird. Then I’d go over it a couple more times with myself, and I’d record it again in an effort to improve fluency.” By re-watching, marking, and commenting, learners could confront the reality of how they speak and could focus their attention on the errors they prioritized.

Some of the feedback regarding the revision system was surprising. Although learners were tasked with noticing where they paused, restarted their speech (repairs), and pronounced or mispronounced their selected vocabulary items, several indicated that they felt they did not know enough to make grammatical corrections. It is surprising because they were not tasked with correcting their grammar but felt compelled to do so in this assignment anyway. The fact that they feel unable to make grammatical corrections on their own, “because I was just barely starting Arabic usually I wouldn’t catch mistakes that I had made...and if there are any mistakes, I would’ve made them when I was writing the draft like for grammar-wise,” suggests that some revisionary tasks may be best left to a TA, whereas others may be usefully done by the learners themselves.

Although one of the major purposes of the recording assignments was to train learners to improve their speaking skills on their own, thus saving teacher time, one of the common pieces of feedback was the importance of TA feedback and a desire for even more of it. One

interviewee explained that the main difference between TA feedback and self-feedback is that the TA can teach the learner about broad principles and how they apply to specific instances in the recording, whereas the learners doing self-feedback may or may not be able to isolate specific errors and will have a harder time relating their errors to broader principles. He added that, while he was grateful for the feedback, he would like to see a balance between error correction and positive reinforcement. “Things like that help me to have confidence. Because a lot of times a TA will look through a video and the things that are good they won’t mark.”

This particular interviewee spent a great deal of his interview talking about TA feedback, pointing out aspects of the design that impacted his learning and opportunities to improve the format of the feedback as well as the content. While the written feedback was useful to him, he felt that audio/video feedback would be especially valuable in helping him to improve his pronunciation. Where the models provide native pronunciation, audio TA feedback could provide a more accessible, intermediate step towards native-like pronunciation: “if that word is pronounced slowly...then I can move closer to the native pronunciation.” Another interviewee echoed this opinion, stating that the audio feedback would be more beneficial to pronunciation than the written feedback. “I mean, you can read how it’s supposed to be pronounced like how it’s spelled in English letters, but you can still say it totally wrong.”

The switch from written to audio feedback is possible within the GoReact app. However, one interviewee suggested moving beyond the app to in-person feedback. For her, real-time, interactive feedback would have helped her to improve more.

If I created the first draft and then somebody walked through my issues, that would’ve been helpful. Because, there’s only so much I can pick out while I’m doing it. And even

with the feedback that I give myself and that I get from the TA comments, it's — I don't know. I feel like I could've had a lot more improvement if I was working with somebody.

Training. Although learners were given an explanation, demonstration, and a preparatory recording assignment as part of their training, interviewees indicated that they were not sufficiently prepared for the assignments. One interviewee explained that although the assignments and instructions were well organized, there was so much for her to process and there was “just a lot of stuff going on.” Another indicated that a tutorial would have improved her experience. In fact, the preparatory assignment in the pilot study included a series of help pages and videos hyperlinked to key areas of the instructions. However, due to the constraints of the university's LMS in the second semester of the study, we were not able to include these hyperlinked help files. One recommendation made by an interviewee was to provide a second demonstration of the assignment process once learners had already begun their recordings. Although the first demonstration was valuable, having an opportunity to experience it for themselves would make a second demonstration more meaningful for them. “I thought I had read everything well, but it still— that went over my head.”

Surveys

Following the interviews, critical incidents of the video recording assignments were extracted from the interview transcripts and categorized. I then selected certain incidents, influences and aspects for use in a survey according to the description in item seven of the trustworthiness section above. All learners enrolled in the Arabic courses were invited to participate in the survey.

Rankings. Respondents were asked to select the five incidents of the video recording assignments that they considered the most critical factors in their learning experience, whether

positively or negatively. They then rank-ordered them from most important to least important of the five. I then selected the top five incidents based on the total frequency of respondent selection and based on a simple medal system, awarding from one to five points in one-point increments, adding all point values for each incident. These rankings are displayed in Table 2.

Table 2

Rankings of Top Five Critical Incidents, Influences, and Aspects of Video Recordings

Rank	Incident	Frequency	Points
1	Content of TA feedback (including specificity)	7	29
2	Doing error correction based on the TA's feedback	9	27
3	Doing your own error correction using your level of Arabic knowledge (i.e., without TA feedback)	7	25
4	Amount of practice learning the assignment before doing it for credit	5	20
5	How well you understood the video models and your ability to extract the vocabulary chunks	6	17

These rankings were only partially in line with the total frequency of citation in the interviews. Three of the items in Table 2, ranks 2, 3, and 4, were also the most frequently cited by interviewees, at 80%, 90%, and 90% respectively (please refer back to Table 1, above). One difference is that “prioritizing task completion over quality” was cited by 70% of the interviewees but only received 8 points. On the other hand, “content of TA feedback” was awarded the most points, but was only cited by 60% of the interviewees. Based on my impression from the interviews, I expected “alignment of the [model] videos and prompts to make the top five list. However, it was only cited by 50% of the interviewees, was selected by only two respondents, and was awarded 6 points.

Follow-up feedback. As part of the survey, respondents provided an explanation for their selections and rankings. In this section, I will discuss key feedback for the five top-ranked items.

1. *Content of TA feedback (including specificity).* Error correction was the most frequently mentioned aspect of the TA feedback in the survey responses. “Specific feedback from the TA is absolutely the most valuable. I’ll never be able to correct myself unless a TA has corrected me first.” Two comments cited the importance of specific feedback rather than generic feedback “The more specific the feedback, the better. This helped me to focus on implementing exactly what was recommended as I created the second draft.” Although one interviewee stated that he would like some of the feedback to focus on things learners said correctly rather than a complete focus on error correction, none of the survey responses brought this up, with the possible exception of “the feedback really helps when it is constructive.”
2. *Doing error correction based on the TA’s feedback.* Although this was the second highest ranked item, most of the feedback for ‘content of TA feedback’ actually related to this item, potentially making it more important than the content itself. Respondents noted that the TA’s knowledge of Arabic made it possible to give useful corrections that the learners could confidently act on. One respondent, who remained in the course only while he was assigned to the self-feedback group, commented that “I would do error correction, but I never actually knew if I was correcting it right. I would’ve benefited greatly from someone there showing me how to correct my Arabic.”
3. *Doing your own error correction using your level of Arabic knowledge (i.e., without TA feedback).* Although feedback on the top two items were generally positive, this item

received mixed feedback. For one respondent, it was critical in that it was harder to improve their second draft without TA feedback. “I was not invested in correcting my recordings, mostly because I did not have the knowledge/experience with Arabic to properly judge my own work.” For others, the challenge of doing error correction on one’s own led to increased linguistic self-awareness and problem-solving skills. As one respondent stated, “doing my own error correction made me more conscious of patterns in my mistakes and helped me more efficiently concentrate on areas of Arabic I could improve on.” Another expressed that “correcting oneself is important, because you remember the thought process that brought you to the conclusion, making the language a multi-use formula instead of a single use answer.”

4. *Amount of practice learning the assignment before doing it for credit.* Practice took different forms for the respondents who selected this item as a critical incident. One respondent described their practice time as mentally constructing their script, then going through it prior to recording. For another, practice involved “[having] enough vocab to really get into a topic.” While practice served learners in their immediate goal of speech performance during their recording, one respondent suggested there was a longer-term value: “when I practiced beforehand I felt like I got the phrases down and could use them later.”
5. *How well you understood the video models and your ability to extract the vocabulary chunks.* Video comprehension was cited as a challenge by most of the respondents who selected it, one went so far as to analogize it thus: “it is similar to preparing a meal for dinner and then not eating the meal. Unless there is a clear understanding of the video, actually partaking of the meal, there is no point in having a demonstration. There is no

point in making the meal if you are not going to eat it.” Another complaint was that the model videos used words that were not taught in class activities or vocabulary lists. “It was hard to extract portions that I didn't understand.” However, other respondents found the videos helpful in creating their own recordings. One of which said:

One of the best ways to learn a language is not theory, but through experience. This is how we naturally learn, and by listening to fluent speakers, it is easier to grasp concepts of vocabulary use because certain things just start to "sound right." The best way to enhance this effect is through listening then doing your own narration based on the patterns in the example.

Conclusions

The purpose of this study was to explore the factors that led to learner success and challenges in a series of video recording assignments based on learner feedback to improve the design or redesign of similar activities. Many critical incidents of the assignments were discovered to have an effect on the learners' experience, with the revision process (especially TA feedback), practice, and comprehension of the model videos ranked as the most influential.

Designers and instructors intending to use asynchronous video draft recording assignments as a means of increasing learner linguistic self-awareness and spoken fluency should consider several things when designing their assignments. Although each of the incidents and categories discovered in this study are useful considerations, the following are key recommendations:

Determine the Focus, Content, and Format of Feedback Given to the Learners

Learners benefit from specific expert feedback (Ericsson, Krampe, & Tesch-Römer, 1993, p. 367), especially in areas where the learners are weak. For instance, although learners

can be trained to notice their pausing and fluency issues somewhat well, they have a harder time correcting grammatical issues. For pronunciation guidance, learners can benefit from a combination of transcribing their own pronunciation and receiving oral feedback and modeling from an expert. This does not come as a surprise, as asynchronous video recording activities have been used by others (Kirkgöz, 2011; Lepore, 2014) to help learners develop their pronunciation skills.

When instructor time is scarce, asynchronous feedback is a viable means of helping learners improve. Griffiths and Graham (2010) supported this claim in a study of three instructional psychology and technology classes in which both learners and instructors felt a sense of instructor immediacy when using asynchronous videos as a means of providing feedback. However, although learners in the present study found asynchronous feedback valuable, some strongly preferred the notion of interactive feedback opportunities when possible.

Provide Learners with Clear Linguistic Models

When the focus is on mimicry, audio and video models should use a limited amount of unfamiliar material so that learners may focus on extraction and incorporation of vocabulary items rather than on listening comprehension. Murphy (2014), in his study on the use of language models to improve English language learner pronunciation, agreed with the importance of model comprehensibility in pronunciation teaching. Because of the difficulty of finding suitable models, providing additional scaffolding may be necessary to allow learners to turn their focus from comprehension to mimicry. Murphy's (2014, p. 265) task list for pronunciation learners is an example of this type of scaffolding.

Provide a Variety of Training Types at Multiple Points

Explanation and demonstration are important for helping learners carry out complex instructions. However, providing opportunities for them to practice the process prior to performing will go a long way in helping them succeed. Additionally, providing readily-available online support in the form of hyperlinked walkthrough videos and explanations will also be of benefit to learners. In support of the notion of providing learners with training, Spanish language learners in Castañeda and Rodríguez-González's (2011) study benefited from creating an ungraded training video prior to attempting their graded work. This allowed them the opportunity to become familiar with the hardware and software in a low risk setting.

Set up the Assignments in Ways that Encourage Practice

This can be done directly through focusing grading on fluency aspects of their recordings or indirectly through focusing instructor feedback on learner fluency. Likewise, requiring the submission of drafts inherently promotes a certain level of practice (Castañeda and Rodríguez-González, 2011, p. 494). It may also benefit learners to make them aware of various methods of practice, for instance: working with a partner, recording practice attempts, or practicing the script in sections.

Create Realistic, Engaging Prompts that Align Closely with Course Content

Although matching the assignment prompts with their respective unit theme is an important baseline, matching it to video models or previously practiced linguistic tasks may be even more beneficial to learners. Where the purpose of the assignment is fluency, it is not necessary to explore new linguistic terrain, and doing so may, in fact, be detrimental to learner fluency development. Prompts used in this study could be improved (a) through asking more provocative questions; (b) by making the learner feel they are being directly addressed, such as

through a conversational writing style; or (c) by providing specific follow-up questions to help learners know how to respond to the prompt. In their study, Castañeda and Rodríguez-González' (2011) provided their learners with such a prompt by addressing it directly to the learners, including a realistic problem, and providing a follow-up prompt in a subsequent assignment worded with greater urgency:

Since you did not renew your lease for your campus dorm, time is running out and you must find an apartment right away. You learned from your past experience. Call the landlord again. Keep in mind that time is running out and you need an apartment and you will be selected by the information you leave in the message (p. 490).

As with all instructional designs, the design of video recording assignments and self-evaluation activities comes down to a series of design decisions. Each decision forces the designer to weigh the importance of effectiveness against practicality. For instance, in this study we found, on the one hand, that asynchronous video recording activities can help save instructor time, but that learners feel they would improve their speaking skills more through direct interactive feedback sessions with their instructor. Instructional designers should carefully consider the key recommendations listed in this section, and should also consider the other findings listed in this study.

References

- ACTFL. (2012). *ACTFL proficiency guidelines*. Retrieved on July 19, 2017 from <https://www.actfl.org/publications/guidelines-and-manuals/actfl-proficiency-guidelines-2012>.
- Blake, R. J. (2013). *Brave new digital classroom* (2nd ed.). Washington, D.C.: Georgetown University Press.
- Butterfield, L. D., Borgen, W. A., Amundson, N. E., & Maglio, A.S. T. (2005). Fifty years of the critical incident technique: 1954 – 2004 and beyond. *Qualitative Research*, 5(4), 475–497. Retrieved from <http://doi.org/10.1177/1468794105056924>
- Butterfield, L. D., Borgen, W. A., Maglio, A.-S. T., & Amundson, N. E. (2009). Using the enhanced critical incident technique in counselling psychology research. *Canadian Journal of Counselling*, 43(4), 265–282.
- Castañeda, M., & Rodríguez-González, E. (2011). L2 speaking self-ability perceptions through multiple video speech drafts. *Hispania*, 94(3), 483–501.
- Clark, R. E. (1994). Media will never influence learning. *Educational Technology Research and Development*, 42(2), 21–29. Retrieved from <http://doi.org/10.1152/advan.00094.2010>
- Ericsson, K. A., Krampe, R. T., & Tesch-Römer, C. (1993). The role of deliberate practice in the acquisition of expert performance. *Psychological Review*, 100(3), 363–406. Retrieved from <http://doi.org/10.1037/0033-295X.100.3.363>
- Flanagan, J. C. (1954). The critical incident technique. *Psychological Bulletin*, 51(4), 327–358.
- Griffiths, M. E., & Graham, C. R. (2010). Using asynchronous video to achieve instructor immediacy and closeness in online classes: Experiences from three cases. *International Journal on ELearning*, 9(3), 325–340.

- Hasson, F., Keeney, S., & McKenna, H. (2000). Research guidelines for the Delphi survey technique. *Journal of Advanced Nursing*, 32(4), 1008–1015. Retrieved from <http://doi.org/10.1046/j.1365-2648.2000.t01-1-01567.x>
- ILR. (n.d.). *ILR speaking skill scale*. Retrieved on July 19, 2017 from <http://www.govtilr.org>
- Kirkgöz, Y. (2011). A blended learning study on implementing video recorded speaking tasks in task-based classroom instruction. *Turkish Online Journal of Educational Technology*, 10(4), 1–13.
- Language Testing International. (n.d.). *How long does it take to become proficient?* Retrieved on July 19, 2017 from <https://www.languagetesting.com>
- Lepore, C. E. (2014). Influencing students' pronunciation and willingness to communicate through interpersonal audio discussions. *Dimension*, 73–96.
- Lin, H. (2015). Computer-mediated communication (CMC) in L2 oral proficiency development: A meta-analysis. *ReCALL*, 27(3), 261–287. Retrieved from <http://doi.org/10.1017/S095834401400041X>
- Murphy, J. M. (2014). Intelligible, comprehensible, non-native models in ESL/EFL pronunciation teaching. *System*, 42(1), 258–269. Retrieved from <http://doi.org/10.1016/j.system.2013.12.007>
- Reed, Y. (2012). Critical pedagogic analysis: An alternative to user feedback for (re)designing distance learning materials for language teachers? *English Teaching: Practice and Critique*, 11(1), 60–81.
- Sugar, W., & Luterbach, K. (2016). Using critical incidents of instructional design and multimedia production activities to investigate instructional designers' current practices

and roles. *Educational Technology Research and Development*, 64(2), 285-312.

Retrieved from <http://doi.org/10.1007/s11423-015-9414-5>

Walker, J. (2015). Perspectives using critical incidents to understand ESL student satisfaction.

TESL Canada Journal, 32(2), 95–111. Retrieved from

<http://doi.org/https://doi.org/10.18806/tesl.v32i2.1210>

Appendix

Categorized and More Inclusive List of Critical Incidents

TA feedback

Online nature of TA feedback

Focus of TA feedback

Timing of TA feedback

Quality of TA feedback

Having the TA's feedback as a model for doing your own feedback

Prompting

Watching native speaker models

Watching a variety of model videos

Authenticity/quality of the prompts

Alignment of the videos and the prompts

Extracting vocabulary words from the videos

Your level of understanding the video models

Course Instruction

Quality of demonstrations

Clarity of assignment instructions

Alignment of video recording assignments and classroom activities

Instructor's classroom feedback on video recording assignments

Placement/location of each step of the assignment

Importance of video recording assignments in class

Consistency of assignment dates

Language Ability and Prior experience

Amount of practice learning the assignment before doing it for credit

Your level of Arabic knowledge

Using words you were more familiar with

Technology

Using GoReact as a recording tool

Ability to listen to/watch yourself

Marking your pauses

Comparing your recording to video models

Transcribing your pronunciation

Recording the response

The number of things you were considering during review/revisions

Having a plan before starting the assignment

Personalizing your response

Practicing your script

Recording your attempts (takes)

Following a verbatim script

Writing a partial script (bullet points or notes)

Feeling pressure to perform while recording/rewatching yourself

Having a personal commitment to the assignments

Doing a 2nd draft

Doing monologs (just you talking by yourself)

Pairing vocabulary with new contexts

Paying attention to your own speech while recording

Other

Importance of Arabic relative to other classes

Amount of time it took to complete video recording assignments

Feeling embarrassed

Amount of time you worked with other students on these assignments in/out of class

Prioritizing task completion (e.g., filling 1 minute of speech) over quality

Timing of when you did first and second drafts

Gaining self-awareness

Ability to work at your own pace

Not reading the instructions thoroughly

DISSERTATION CONCLUSION

Asynchronous video recording activities have been and can be used in numerous ways to help learners improve their speaking skills, whether from the standpoint of communication or specific aspects of their speech production. Such activities range from video pen pal correspondence to the production of informational videos to iterative draft recording to focus on specific types of language gains. In my first article, I conducted a literature review, surveying the current uses and implementation challenges of asynchronous video recording activities, as well as the methods of studying their effects. Of particular note is the fact that current research on the effect of these types of activities tends to focus on accuracy, fluency, and pronunciation as measured by learner perceptions of gains in these domains. Although learner perceptions can provide good insights into learning experiences and learner attitudes, they are not appropriate as measures of linguistic growth.

To further investigate the effects of asynchronous video recording on fluency, I implemented one type of video recording and feedback activity in an Arabic classroom for my second article. Fluency was measured in terms of mean syllable duration, mean length of utterance (in seconds and syllables), and average silent pause duration, each normalized based on L1 fluency measures. Although there were few significant findings, the effect sizes were moderate and noteworthy. Additionally, the fact that there were significant results in view of the small sample size and short treatment period suggests that more significant results could be found in a classroom focusing on these activities.

As a byproduct of studying the effect of these activities on fluency, I also found that latent growth modelling (LGM) was more effective than the more commonly used *t* tests and regression models. This is because learners do not always prioritize and complete class

assignments, leading to incomplete data sets. LGM was able to incorporate data from learners who did not complete one draft of a given recording assignment, or who did not complete either the first or final draft of a treatment type, thus increasing the sample size. Given the realities of using classroom-based research, LGM is a useful means of analysis foreign language learning research.

In my third article, explored the experiences of learners participating in asynchronous video recording and feedback activities. This study revealed many factors that go into the design of video recording activities. Learners were especially interested in receiving interactive feedback instead of recorded asynchronous feedback; in responding to realistic and provocative prompts; in having comprehensible, relevant language models; and in having opportunities to practice before submitting their own recordings. While the focus of this study was on producing specific design recommendations for asynchronous video recording activities, it also produced a list of more general themes that could be used to more broadly describe the learners' experiences.

One of the goals was to find a means of training learners to be more self-sufficient in improving their speaking skills through thoughtful self-evaluation. However, an important finding was that, although learners found the practice in self-evaluation helpful, some learners felt that direct, synchronous interaction with an instructor or teaching assistant would allow them to get more valuable feedback. Thus, while asynchronous video recording activities may alleviate some of the demands on teacher time, it should not be used to completely replace synchronous, interactive teacher-learner feedback.

Overall, designers and instructors should consider how asynchronous video recording (whether for communication or for targeting fluency or other aspects) should fit into their

curriculum. On the one hand, it should at least be a design consideration, as it provides learners with additional speaking opportunities outside the classroom as well as a means to critically analyze their own speech. However, its use comes at the cost of other language learning activities, such as reading, writing, and listening, as well as activities focusing on grammatical knowledge and vocabulary building. Ultimately, instructional designers (including instructors) must weigh the opportunity costs and benefits of asynchronous video recording activities against other activities before incorporating or rejecting them. However, because of their potential value, they should at least be a design consideration.