

APPLIED LANGUAGE LEARNING



Applied Language Learning



Applied Language Learning

*Volume 27, Numbers 1 & 2
2017*

Defense Language Institute Foreign Language Center
Presidio of Monterey

Editor

Jiaying Howard

Editorial Board

Alessandro Benati – *University of Greenwich*

Steven Berbeco – *Foreign Service Institute*

Christine M. Campbell – *International Language and Culture Foundation*

John S. Hedgcock - *Middlebury Institute of International Studies at Monterey*

Eli Hinkel - *Seattle Pacific University*

Chuanren Ke - *University of Iowa*

Ryuko Kubota - *University of British Columbia*

James F. Lee - *University of New South Wales*

Margaret Malone - *Center for Applied Linguistics*

Scott G. McGinnis - *Defense Language Institute Foreign Language Center*

Rebecca Oxford - *University of Maryland*

Maria Parker - *Duke University*

Richard Sparks - *College of Mount Saint Joseph*

Susan Steele - *International Language and Culture Foundation*

Hongyin Tao - *University of California, Los Angeles*

Copy Editors

Jeff Hansman

Michael McGuire

Applied Language Learning

PB 65-17-01 & 02

The mission of Professional Bulletin 65, *Applied Language Learning* (US ISSN 1041-679X and ISSN 2164-0912 for the online version), is to provide a forum for the exchange of ideas and information on instructional methods and techniques, curriculum and materials development, assessment of needs within the profession, testing and evaluation, and implications and applications of research from related fields such as linguistics, education, communications, psychology, and the social sciences.

Applied Language Learning, published semiannually by the Defense Language Institute Foreign Language Center and Presidio of Monterey, presents professional information. The views expressed herein are those of the authors, not the Department of Defense or its elements. The content does not necessarily reflect the official US Army position and does not change or supersede any information in official US Army publications. *Applied Language Learning* reserves the right to edit material.

Further reproduction is not advisable. Whenever copyrighted materials are reproduced in this publication, copyright release has ordinarily been obtained only for use in this specific issue. Requests for reprints should be directed to the authors.

Availability

To access *Applied Language Learning* online, go to:

<http://www.dliflc.edu/publications.aspx>

Additionally, you may obtain the journal on microfilm from ERIC Clearinghouse on Language and Linguistics, Center for Applied Linguistics, 1118 22nd Street, NW, Washington, DC 20037.

Postmaster

Send change-of-address information to:

Applied Language Learning
Defense Language Institute Foreign Language Center
Presidio of Monterey, CA 93944-5006

United Parcel Service Customers

Location is:

Applied Language Learning
Bldg. 614, Room 243
Presidio of Monterey, CA 93944-5006

Readers and Authors

Contact Editor, Dr. Howard
(ATFL-ASD-AJ), *Applied Language Learning*
jjaying.howard@dliflc.edu

Printing Coordinators
Webmasters

Tom Colin & Ricky Harris
Natela Cutter & Dusan Tatomirovic

Applied Language Learning

Volume 27, Numbers 1 & 2, 2017

CONTENTS

ARTICLES

- 1 Across the Threshold: A Call for ILR Proficiency Descriptor Banks
Don Holman
- 18 Using Writing to Teach Pronunciation: An Experimental Fourth-Year
University Spanish Phonetics/Phonology Course
Michael D. Hubert & Donny Vigil
- 41 The Order of Explicit Information in Processing Instruction
Erin McNulty Diaz
- 73 Correlating Aptitude with Oral Proficiency: A Quantitative Study of
DLAB and OPI Scores Across Four Language Categories
Isabelle Poupard Santizo

GENERAL INFORMATION

- 97 *ALL* Indexes (2007-2017)
- 102 Upcoming Events 2017
- 104 Information for Contributors
- 107 Call for Papers

THANK YOU REVIEWERS

Applied Language Learning relies on expert reviewers for quality of the journal. Special thanks go to the individuals listed below. The publication of *Applied Language Learning* was made possible with their generous support.

Netta Avineri
Leila Bernardo
Tsengtseng Chang
Martine Danan
Jack Franke
John Hedgcock
Hyunsoo Hur
Chuanren Ke
James Lee
Ruiqi Ma
Scott McGinnis
Maria Parker
Thor Sawin
Hongying Tao
Rong Yuan

Alessandro Benati
Christine Campbell
Marina Cobb
Vera Dumancic
Hyekyung Sung-Frear
Eli Hinkel
Gordon Jackson
Ryuko Kubota
Lisa Leopold
Jason Martel
Rebecca Oxford
Jean Ryu
Susan Steele
Deanna Tovar
Jihua Zhou

Across the Threshold: A Call for ILR Proficiency Descriptor Banks

DON HOLMAN

Defense Language Institute Foreign Language Center

Whereas the Interagency Language Roundtable Language Skill Level Descriptions broke new ground for assessing proficiency in foreign languages, the need for user-oriented (rather than assessment-oriented) proficiency scales has led, especially in Europe, to the creation of scales consisting of positively formulated “can-do” statements, which describe sociolinguistic functions and are calibrated to the proficiency levels of the Common European Framework of Reference for Languages (CEF or CEFR) and other European scales aligned with CEFR. To mitigate the washback effects of ILR-based proficiency assessment, which lead to an all but exclusive focus on threshold linguistic functions, this paper advocates banks of proficiency descriptors calibrated to the ILR base or plus level to which they belong. Such descriptor banks would provide explicit standards for foreign language acquisition in U.S. government agencies and programs and resolve many problems that arise in foreign language acquisition from defining general language proficiency exclusively in terms of threshold-level linguistic skills.

Keywords: *foreign language assessment, can-do statements, Common European Framework of Reference for Languages, threshold systems, gestalt concept, general language proficiency*

LANGUAGE PROFICIENCY SCALES

The year 2001, while linked in American memory with one dark day in September, commenced with a hopeful outlook for language educators in Europe. By autumn of that year, the Council of Europe had prepared a publication that was to facilitate an era of heightened communication among the nations of Europe. The *Common European Framework of Reference for*

Languages (CEF or CEFR) introduced not only a hierarchy of proficiency levels comparable to, if distinct from, the Interagency Language Roundtable (ILR) Language Skill Level Descriptions,¹ it also offered a vision of *plurilingual* Europeans, who would possess the linguistic and intercultural skills to overcome centuries-old language and ethnic barriers and work together in behalf of European prosperity (See Council of Europe, 2001, pp. 4-5). Although not the most significant initiative of the European Council, the CEFR stands as a declaration of European independence from the ILR Descriptions, which for decades defined foreign language proficiency for the U.S. and its NATO partners.

As important as the CEFR has become for foreign language acquisition and assessment within the European Community, the two decades that preceded its publication saw the introduction and ever wider application of scales for measuring language proficiency. North and Schneider (1998) note in particular “a proliferation of European scales” since 1980, which were not derived from the ILR Language Skill Level Descriptions (2016, pp. 217-218). The multiplicity of language proficiency scales reflects the divergent motives that institutions have for establishing criteria or standards and the uses to which they are applied.

According to Bachman (1990), factors that influence and inform different types of foreign language measurement include the ways language ability may be defined, the approaches taken to assessment, and the contexts that call for measuring skill in a foreign language. Bachman (1990) distinguishes, for example, the way proficiency is described in the ILR, American Council on the Teaching of Foreign Languages (ACTFL), and related scales on the one hand, and the definition of communicative ability that he advocates.² He contrasts the *real-life* approach of the ILR and ACTFL scales, which assign a global score to describe an individual’s functional language skill, with the *interactional/ability* approach that rates key features of the examinee’s sociolinguistic competence separately (pp. 325-330). Whereas for Bachman (1990) scales and assessment are largely synonymous, Alderson (1991) attributes three important functions to proficiency scales in particular, depending on the purposes for which they are created and applied. He characterizes scales as user-oriented, assessor-oriented, or constructor-oriented, depending on the extent to which they are intended, respectively, for “reporting results, guiding the rating process, and controlling test construction” (p. 74). The ILR Descriptions, for example, may be used by an agency to predict what jobs a team member’s ILR score implies she should be able to perform (user orientation); by oral proficiency testers as guidance for eliciting and rating an examinee’s speech (assessor orientation); or by curriculum developers to write appropriate textbooks and unit assessments for a basic or advanced language course (constructor orientation). Alderson (1991) illustrates these functions of proficiency scales with examples of tests and test methods that were either created or revised to render them consistent with scale band descriptors, which provided clear points of reference and guidance for test development.

North and Schneider (1998) review and elaborate criticisms that have been raised about the global band or level descriptions that are typical of proficiency scales (pp. 219-221). Concerns include the generally intuitive nature of most scale level descriptions, lacking basis in an explicit theory of measurement, and a general tendency to infer scale validity exclusively from high rates of inter-rater reliability, often itself the result of extensive training and experience with the scale. Another issue that has arisen from widespread use of proficiency scales is the disproportionate importance of the borderlines that separate adjacent bands or levels. As teachers and students in high stakes foreign language programs endeavor to raise communicative ability beyond the learner's starting skill level, they typically focus attention on the competencies laid out as definitive in a scale's description of the next higher target level. Whereas scale bands are usually conceived wide enough to encompass a variety of individual language users, questions and ambiguities inevitably arise over performances that either barely cross or fall just short of a borderline. In training programs where this occurs often, testers may develop supplementary measurement tools, such as rating grids that elaborate borderline criteria (See Lowe, 1982, pp. 3-7), or even abandon broad proficiency level definitions in favor of what Upshur and Turner (1995) term *empirically derived, binary-choice, boundary-definition* (EBB) scales: which are developed exclusively to measure whether an examinee has crossed the borderline from one level to another (p. 6).

Even the proliferation of proficiency scales that proclaim their independence from the ILR Language Skill Level Descriptions attests to the importance of the ILR in their particular hierarchy of levels. Like the ILR, the level descriptions in most scales exhibit a dual interest: to specify ability levels with sufficient detail to demarcate the differences from one level to another, as well as to generalize the bands so that each encompasses a wide range of individual abilities that have common features. This legacy of the ILR Descriptions has arisen from decades of utility that government agencies and national education bodies have found in classifying the competencies of second language users according to a single, common system of measurement. Nevertheless, the satisfaction of the creators and users of the ILR Descriptions over its stature and success should be tempered by consideration of the motives and choices that informed the development of alternatives to the ILR scale. The Common European Framework of Reference, as the most widely used scale other than the ILR, merits special consideration, due both to what the ILR and CEFR have in common and to what sets the CEFR apart from its American predecessor and model.

Perhaps the most striking difference is the relative brevity and simplicity of the CEFR's primary level definitions compared to the longer and more detailed ILR Language Skill Level Descriptions. Instead, the CEFR collaborators opted to supplement its "calibrated scale of language proficiency" with "a bank of classified, calibrated descriptors" (North and Schneider, 1998, p. 218). These long lists of stand-alone descriptors—gleaned mostly from other, related scales and sub-divided by level, skill, and other sociolinguistic

categories—provide rich detail missing in the *global scale* of the CEFR's Common Reference Levels (see Council of Europe, 2001, p. 24).³ Yet for ILR users in agencies of the U.S. Government (USG), this raises issues that revolve around one central question. Does the new and more detailed alternative, the CEFR, necessarily result in greater utility or validity compared to the ILR—which has served so long and so well the purpose for which it was designed? The interest of this study is twofold: 1) to examine the ILR approach to measuring foreign language proficiency, including the history of its development and application, in order to identify the experiences and decisions that have shaped it as a system of assessment; and 2) to determine if alternative scales such as the CEFR offer models for addressing issues in ILR based proficiency testing.

GOVERNMENT AND ACADEME: THE DIVIDE IN FOREIGN LANGUAGE EDUCATION

The proliferation in the 1950s and 1960s of college foreign language programs, notably for French, German, and (Iberian) Spanish, both reflected and reinforced a national preoccupation in the U.S. with the peoples and languages that shared with English the common cultural heritage of Western civilization: the defense of which was then led by the United States. During this period, preoccupation at universities with preparing students to read and value the great works of literature typically relegated courses devoted to language acquisition to the first two undergraduate years.⁴ As often as not, however, these programs produced graduates with a spotty knowledge of a nation's literature and still spottier ability to communicate in the language.

By contrast, the success of language training at U.S. government schools, such as the Foreign Service Institute (FSI), which produced graduates after a few weeks or months who were proficient at one ILR level or another in the language they recently had learned, attracted the attention and criticism of academe: which had been neither consulted nor kept abreast, through published research, of developments at federal agencies that had contributed to this success. By the late 1980s, following publication of the ACTFL Proficiency Guidelines, which adapted the ILR Skill Level Descriptions to the needs of learners enrolled in public schools and college programs, American scholars raised questions about the USG and ACTFL testing practices (see e.g. Bachman and Savignon, 1986). Their scholarly, if belated, interest prompted Lowe at the Central Intelligence Agency (CIA) Office of Training and Education, to publish a brief history and overview of language proficiency testing derived from the ILR scale.

Apart from a short chronology of the ILR and ACTFL scales and a catalog of testing practices and guidelines, Lowe's "Unassimilated History" (1988) accounts for the ILR-based testing methodology and its soundness, on the one hand, and explains why these developments had proceeded largely unnoticed, or rather *unassimilated*, by university scholars in languages and

linguistics. Lowe admits that university-trained ILR collaborators did little to solicit the attention or acumen of researchers in the relevant fields, noting but a “single early publication of proficiency testing” at the FSI that appeared in 1959: the one exception to a 25-year silence (See Rice, 1959). He stresses that “little notice” was taken “in academia” of the 1959 article and suggests that this scholarly disinterest in government testing was one reason “for the small number of studies available” at that time on the USG language testing (p. 16).

In all fairness to academe, the 1959 article was a one-page report, rather than a study, and appeared in the second issue of a new periodical—*Linguistic Reporter*, the title, format, and contents of which were suggestive of a newsletter rather than a peer-reviewed journal. Lowe (1988) concedes, for that matter, that the complaints of the scholarly community—“that the ILR conducted little research, rarely published it,” and failed to coordinate its work with public and higher education—were “generally accurate [...] from the 1950s until 1974” (p. 25). It is to Lowe’s credit that none contributed more intelligently and articulately than he in the following decades, to initiate scholarly dialog and partnerships with academe. Bachman (1990) even calls Lowe “one of the major spokespersons” for ILR proficiency testing (p. 5). But can the same be said of his predecessors? What is to be made of the all but silent period from 1954 to 1978? Did their work lack a scholarly foundation?

We need in fact look no farther than the same two articles by Rice (1959) and Lowe (1988) to find a reason for limited scholarly outreach from USG proficiency testing. Rice explains in his article the test procedures and assessment metric applied by the Testing Unit at the FSI in the speaking and reading proficiency interviews. He highlights insights won from statistical analysis of ratings and notes changes introduced as a consequence of those insights. His outline of rating methods at the FSI is intriguing and might indeed have deserved notice from university researchers. But Rice writes that the “staff of the Testing Unit feels that it is too early to publish” its findings; that the samples “yielded somewhat different results”; and “that a larger and more representative sample is necessary” (p. 4). Rather than scholarly negligence, Rice’s account evinces scientific scrupulousness in the university-educated testing staff, who were hesitant to publish claims based on less than conclusive data.

Nevertheless, although this accounts for some delay, it does not explain why linguists at the FSI published no scholarly research at all for several years. Lowe supplies a possible answer. He contends that testers at the USG language schools never intended to invent their own proficiency testing system and would have been prepared to adopt or adapt one from academe. But as college language departments were at the time “focused heavily on literature” rather than on speaking competence, government schools, having “the requisite expertise” among their staff, proceeded to go about meeting their own needs (p. 16). The general picture Lowe paints is one of hands-on scholars simultaneously engaged in cutting-edge research and development in an environment of urgent needs and time constraints.⁵ The reticence of the FSI testing staff about

publishing their data reflects a busy schoolhouse that perpetually fails to satisfy laboratory research conditions; where the latest statistics (still inconclusive) prompt yet finer revisions in test procedures and rating metrics almost before the ink with which the data were recorded has dried.

A term that Lowe uses more than twenty times in the 1988 article to refer to the ILR proficiency scale, including testing procedures related to it, is his characterization of these collectively as a *system*. Not only the ILR *definitions*—the word he prefers when referring to the Skill Level Descriptions—but the *ILR system* of proficiency assessment is what had yet to be *assimilated* by academia in the 1970s. The diverging scholarly interests and professional practices of university researchers versus those of USG linguists, perceptible already in 1959, had led by 1974 to nearly total separation: university researchers in linguistics, with their scientific method of inquiry and growing body of scholarly literature; and linguists in government service, with their proficiency testing *system* and its demonstrable validity. To the system's merits one should add examinees whose measurable proficiency was acquired in the USG language schools that were held to the system's rigorous standards. It is the success of the ILR system, and especially the proficiency-oriented language training it engendered, that in the 1970s attracted the attention of representatives from public and higher education, who were dissatisfied with the low success rate of high school and college foreign language instruction. A partnership of the Interagency Language Roundtable, the Educational Testing Service, and the American Council on the Teaching of Foreign Languages led to publication of the ACTFL Proficiency Guidelines, whose exclusion of Levels 4 and 5 and more detailed elaboration of the lower levels made the ILR scale more relevant to academe.

By the 1980s, it might be said that the ILR system included not only the body of proficiency testing methodology based on the scale but also, in an extended sense, well established language schools within government agencies: with organizational structures resembling those of universities and many language faculty possessing higher degrees. All of these organizations, their professional standards, curricula, and tests, were aligned with the ILR system of proficiency assessment. Be that as it may, Lowe is undoubtedly **not** thinking of foreign language instructional programs when he chooses the word *system*. On the contrary, he writes at some length about the misunderstanding—overlayering—that arises in proficiency testing “when statements describing *learner* behaviors are superimposed on earlier [ILR] statements describing *user* behaviors” (Lowe, 1988, p. 21). In other words, the validity of the ILR scale is compromised whenever its proficiency descriptors are interpreted as final learning outcomes for those who recently acquired a language in a course of instruction. He stresses that the level descriptions refer to “users” whose proficiency level is already well established rather than to learners whose language is still under construction and not yet fully internalized through regular use. Moreover, “confusion” over who is described in the ILR descriptions was “exacerbated by academia’s burgeoning interest in proficiency based curricula,

emphasizing learners,” whereas “the [ILR level] definitions were never intended for this population” (Lowe, 1988, p. 21).

Lowe’s concerns about the danger that washback poses to test validity are understandable; although his particular worry is over tester rating practices in a school setting rather than faculty and students teaching and learning to the test. In their “Introduction” to the journal volume that features the 1988 article, Lowe and Stansfield (1988) point to “a surge of interest in second language proficiency assessment within the academic community” (p. 1). As true as that may have been, they might have named a related phenomenon by omitting just one word, namely: a surge of interest in second language proficiency within the academic community. The *proficiency movement* to which Lowe (1988) refers in his opening sentence encompassed more than just the definition of proficiency in the ILR and ACTFL scales, but included and, for most academics, lay precisely in the approaches to teaching and learning most suitable to cultivating proficiency in a foreign language. Even the title selected for the ACTFL scale is revelatory. The choice to call them “Proficiency Guidelines” rather than descriptions or definitions represents a shift in focus that acknowledges their implicit relevance to teaching and learning, as well as assessment. Although their importance for language acquisition gets no mention until the 1999 edition, the preface to the 2012 ACTFL Proficiency Guidelines acknowledges that they “have had an increasingly profound impact on teaching and learning in the United States” (ACTFL, 2012a, p. 2).

THE “ILR SYSTEM”: OF THRESHOLDS AND GESTALTS

We have shown that Lowe’s characterization of the ILR as a system refers exclusively to its role in assessing foreign language proficiency, whereas it does not include the approaches to language acquisition that result in proficiency. To determine the reason for his preference of the term *system*, as well as its semantic range, we can find a telling reference in the training manual he authored, the *ILR Handbook on Oral Interview Testing* (1982). There he characterizes the ILR Skill Level Descriptions and ILR-based testing methodology as a “threshold system”, rather than a “midpoint system”, in the sense that “a constellation of factors [is] clearly present when the major border between two levels is crossed” (pp. 3-10). The thrust of Lowe’s reasoning is that proficient speakers who fall within the same ILR level exhibit, even at a level’s lower limit or *threshold*, a distinctly similar, definitive linguistic profile.

A word Lowe uses to characterize the profile shared by all speakers who fall within an ILR level, is *gestalt*. Whereas the German word, signifying a figure or form, may be used in a variety of contexts, its entry into the American scholarly lexicon was chiefly by way of Gestalt psychology. Summarizing his earlier exposition of the ILR threshold system in the *ILR Handbook*, Lowe (1988) quotes his original definition of a threshold as “a constellation of factors that are clearly present *to perform the function(s) in question* when the major border between levels is crossed” (p. 23, my italics). The words *gestalt*, *figure*

and *constellation* reflect the slogan Lowe (1982) chooses for the third chapter of the *ILR Handbook of Oral Interview Testing*, which is devoted to rating an examinee's language production: "It's wholes, not parts" that "you rate" (pp. 3-4). Rather than just analyze an examinee's language in terms of lexical precision, pronunciation, and grammar (something OPI testers are also trained to do), Lowe stresses that proficiency that meets the threshold criteria for any level constitutes a functional *whole* characteristic of the level, such that the proficiency of one examinee bears a resemblance to that of other speakers at that level. This is not to say that they have all strengths and weaknesses in common; some *parts* of a profile may presumably differ from those of another profile. But following Lowe's argument, we can deduce that something like a generic gestalt is typical of all foreign language users who are proficient at the same level.

Lowe (1988) compares the implications of his gestalt concept in the ILR threshold system versus a midpoint system such as the ACTFL scale, with its low, mid, and high subdivisions of the lower levels. Whereas a "weak gestalt" is present at the lower sub-range of a level in a midpoint system, becoming "strong" in the mid-range: the functional gestalt in a threshold system is already strong; i.e., clearly evident, once the speaker is beyond the lower borderline of a level (p. 23). For Lowe (1988), "the *gestalt* nature of the ILR system" is virtually synonymous with its threshold nature (p. 23). The "*gestalt* nature of the ILR" is also a reason why even those new to the ILR often have a clear idea, once they are familiar with the Skill Level Descriptions, which kinds of foreign language speakers are described. That is, the features or factors enumerated at each level describe such speakers well enough that readers recognize who is described.

The problem arises, however, that when ILR neophytes rate individual speakers, they often disagree in their ratings. That is to say, the "perceptual gestalt" is—as is so often true of perception generally—in the eye of the beholder. In fact, the oral interview score sheet included in Chapter 3 of the *ILR Handbook* (Lowe, 1982) includes categories for vocabulary, grammar, pronunciation, sociolinguistic competence, as well as the level-specific tasks used to elicit the language to be rated. A rater scores each of these independently in terms of the level to which the examinee's performance corresponds for that rating sub-category. According to Rice (1959), rating each factor separately resulted in "a high degree of consistency in the subjective judgments of the examiners." Whereas Lowe (1982) subordinated the *parts* to the *whole*, analyzing the *whole* of an examinee's language in terms of its constituent *parts* had resulted in a higher rate of reliability among examiners at FSI. Similar considerations undoubtedly motivated Lowe's (1988) caution that the ILR "definitions cannot replace hands-on exposure"; that "[p]erhaps the best access to the system is offered by first-hand experience: observing tests, being trained as a tester, or being trained as a tester trainer" (p. 20). The testing and rating practices that oral proficiency testers employ, in their watch over the thresholds and plus-level transitions, cannot be inferred from the ILR Skill Level Descriptions alone.

It is the threshold nature of the ILR system, too, that has come to dictate the communicative *tasks* that testers elicit from examinees during the oral interview. Given the categorical importance of the base level thresholds and plus level transitions, the speaking tasks that examinees are asked to perform are limited to those whose difficulty is calibrated to the lower threshold of each level. Moreover, the ILR Skill Level Descriptions (2016) dictate that a plus level is awarded only “when proficiency substantially exceeds one base skill level and does not fully meet the criteria for the next ‘base level’.” In other words, it is the degree of success or failure when performing tasks at the level immediately above that dictates whether an examinee is awarded a (lower) plus level. From the standpoint of assessment then, the threshold nature of the ILR system defines and measures proficiency exclusively in terms of threshold-level communication, because assessment specialists are “focusing on thresholds in interpreting the scale” (Lowe, 1988, p. 24).

THRESHOLDS, RANGES, AND PROFICIENCY PROFILES

Nevertheless, experienced testing specialists are apt to stress that ILR levels comprise expanding *ranges* of communicative ability that encompass ever more, progressively complex linguistic functions, demanding social contexts, and rigorous accuracy requirements as one moves up the scale. For this reason, too, even examinees who exhibit marked differences in their respective ability to communicate may still fall within the same level. This is always the case, as long as each of them can perform successfully the required threshold tasks for their shared level but cannot (or can but partially) perform the threshold tasks for the next higher level. When proficiency-testing specialists wish to stress the unique pattern of strengths and weaknesses characteristic of a particular language user, they refer to that individual’s *linguistic profile*. Even if one examinee (e.g., a so-called “heritage speaker”) exhibits substantially more facility and native-like expression than another one who, say, acquired the language through classroom instruction, their respective linguistic profiles may still fall within the same level. For those strengths and weaknesses that deviate from the generic gestalt of the typical threshold performance are of less significance for assessment specialists, who are focusing on thresholds.

At this point, a semantic shift in the meaning of the word *profile* is noticeable. Whereas in Lowe (1982, 1988) the word is virtually synonymous with *gestalt*, referring to the generic proficiency that is characteristic of ability at or above an ILR level; testers refer to a speaker’s linguistic profile when they are interested in her unique pattern of strengths and weaknesses. Lowe (1982) points to the lexicon of Gestalt psychology when he calls this stress on uniqueness “sharpening”, in contrast to the “closure” that stresses the similarity of an individual profile and the generic gestalt of a level (pp. 3-9). In either sense of the word, however, assessment specialists are interested in identifying the ILR level to which an individual’s proficiency belongs. Is the examinee at ILR Level 2, 2+ or 3? When one is focusing on thresholds, unique features

recede and may even be disregarded. Even when these features are identified, they are bracketed out, as it were, to simplify rating. In the lexicon of Gestalt psychology, sharpening is minimized in the interest of closure.

Focusing on thresholds in the ILR system of proficiency assessment has implications for language acquisition that fall under the heading of *test washback*. In USG language schools, it is not just testers and examiners, as Lowe (1988) has shown, whose interpretation of the ILR descriptions as learner behaviors *overlays* their original sense. Teachers and learners, too, are apt to focus language acquisition on those linguistic tasks, topics, and accuracy criteria that are used in assessment to elicit and measure the examinee's language production. After all, job assignments, promotions, and raises not infrequently depend on whether a USG employee attains the requisite ILR level in a proficiency assessment. Such high stakes tests are the very pump, as it were, that generates washback. In the worst instances, it can result in all but exclusive emphasis on and rehearsal of the topics, question types, and tasks that are used in the examination to sample and measure the students' knowledge or skill.

In the ILR system, experienced testers and well-designed assessments can usually circumvent such blatant instances of teaching and learning to the test. But another question emerges from the history and methodology of the ILR system: what are the implications of focusing on thresholds in interpreting the [ILR] scale? Another related question that is relevant to language teachers and learners is the following: what should be the language acquisition objectives for learners whose proficiency profile has already assumed the general contours of the threshold gestalt for one level? Should learning be oriented on reinforcement and expansion of linguistic functions already mastered, or on higher-level tasks?

In most USG language programs, the most reasonable goal is assumed to be the next higher plus or base level. Moreover, as oral interview testing assesses plus-level performance by measuring the degree of success or failure in performing higher-level tasks, teachers and learners normally set their sights on the communicative tasks used to test the next higher base level. We recall that the range encompassed by each level implies not only improved lexical and structural precision, but a spectrum of other communicative functions that fall within the sociolinguistic domain encompassed by a level, which are beyond the narrow scope of threshold criteria. Nonetheless, whereas these linguistic competencies fall implicitly within the sociolinguistic scope of a level, they are not formulated explicitly in the skill level description, because they lie beyond the threshold that is the focus of the description. As a consequence, relatively little thought or instruction is devoted in USG language schools to competencies absent from the ILR Descriptions, compared to the time spent mastering the tasks and content that are known (or assumed) to be used by testers or in tests that assess threshold proficiency.

“CAN-DO” DESCRIPTORS AS PROFICIENCY BENCHMARKS

Apart from its levels, the thresholds or cut-off points of which are calibrated differently than the ILR skill levels, the *Common European Framework of Reference* was, from its conception, intended to offer guidance as much for teaching, learning, and curriculum design as for assessment. For this reason, the primary skill level descriptions consist exclusively of “can-do” statements and include no references to characteristic limitations (i.e., “can’t-do” descriptors). This is what the CEFR authors mean by its “action-oriented approach”, which describes language proficiency in terms of the “language activities” carried out by “social agents” (Council of Europe, 2001, p. 9).

Nevertheless, the level descriptions of the “Global Scale” are quite minimal compared to the ILR descriptions. That of level B2, for example, also known as Vantage level, consists of just three descriptors:

Can understand the main ideas of complex text on both concrete and abstract topics, including technical discussions in his/her field of specialisation. Can interact with a degree of fluency and spontaneity that makes regular interaction with native speakers quite possible without strain for either party. Can produce clear, detailed text on a wide range of subjects and explain a viewpoint on a topical issue giving the advantages and disadvantages of various options. (Council of Europe, 2001, p. 24)

Such a brief description, omitting references to limitations or weaknesses, provides sketchy specification, leaving considerable doubt over the gestalt that is typical of the level, as well as where to mark the thresholds between Vantage and the levels above and below it. Moreover, the descriptors include references to receptive and productive skills, as if CEFR assumed that all language users are equally proficient in all skills.

In fact, the CEFR Global Scale is supplemented with both skill-specific descriptions for each level and level-specific accuracy statements. Moreover, the original publication, the *Common European Frame of Reference for Languages: Learning, Teaching, Assessment* (Council of Europe, 2001), includes a chapter of “Assessment” that stresses the need for comparative grids for rating, which address both expected communicative functions and accuracy descriptions for vocabulary, grammar, etc., at each level. Nonetheless, the specific objectives, in terms of linguistic functions and accuracy requirements, are left for the users—agencies, businesses, schools—to choose, permitting them to focus on the sociolinguistic actions (i.e. functions, tasks) that are relevant for their purposes.

Indeed, even though proficiency descriptions of the Global Scale provide a standard in the form anchor descriptors—a fixed point of reference for each level, the CEFR assumes that a global level description is but the most general taxonomic classification. A company that needs a purchasing manager for a particular linguistic region will want a Vantage level speaker who can

negotiate over production deadlines and pricing, rather than another Vantage level speaker who can interview and advise refugees from a war-torn region. To the extent that such skills are to be the outcomes of instructional programs, considerations for course design must include the prerequisite proficiency level for learners, the target proficiency level, integration in the curriculum of global sociolinguistic functions and job-relevant communicative skills, and measurement standards for assessment.

Rather than leave it to the CEFR users to formulate language acquisition objectives, European partner organizations interested in proficiency standards have created descriptor *banks*—a word that denotes their openness to growth; i.e., to additional descriptors—which are comprised of can-do statements, organized by linguistic skill, sociolinguistic function, and proficiency level, describing what an individual “can do” to communicate in a foreign language.⁶ The trend to create scales consisting entirely of positively formulated descriptors is intended to make the scales *user-oriented* rather than just *assessor-oriented*, like the descriptor lists made available on the Interagency Language Roundtable website for the purpose of *Self-Assessment*. Instead of *focusing on thresholds*, such descriptor banks provide a detailed perspective on the variety of communicative functions that lie across a proficiency range. Whereas in a threshold system such as the ILR, teachers and learners tend to focus on the next higher threshold to orient language acquisition beyond one’s present skill level, descriptor banks represent an inventory of sociolinguistic functions that are at a user’s level of emerging or expanding ability, and which may be attainable through lexical/contextual retooling of established functions, correcting fossilized errors, or simply learning what has been omitted at one’s present level of development.

Comparable efforts would seem to be taking place on this side of the Atlantic, where ACTFL has redoubled its activities to articulate foreign language standards, based on the ACTFL Proficiency Guidelines, for schools (K-12) and undergraduate college education (13-16). Recent fruits of its work include the ACTFL (2012b) *Performance Descriptors for Language Learners*, and the *Can-do Statements*, published jointly by the ACTFL and the National Council of State Supervisors for Languages (2014). The *Can-do Statements* in particular “mirror much of what is addressed in the Common Core State Standards (CCSS) for English Language Arts (ELA) and Literacy” (NCSSL and ACTFL, 2014, p. 2). For that reason, the descriptors are most numerous for the lower, developmental levels (Novice through Advanced). As Americans rethink, then, the knowledge and skills essential to prepare our youth for competition in a global economy, the work of the ACTFL and its partners only makes clearer a comparable need to map out, in much greater detail, the linguistic landscape of the workshop, office, and boardroom: the sociolinguistic domains of ILR levels 2+ through 5.

With its recent publication of level descriptions for translation, audio translation, interpretation, and intercultural communication, the Interagency Language Roundtable has responded to a strategic need for language and

intercultural skills beyond the general proficiency outlined in the ILR Language Skill Level Descriptions. As for the creation of descriptor banks, as this paper advocates, the initiative should perhaps come from the ILR member agencies, who undoubtedly know best which linguistic skills are especially relevant for their associates. Nevertheless, the expertise of the Roundtable, based on half a century of experience measuring foreign language competence, might provide essential direction, coordination and oversight, to ensure that descriptors are neither redundant nor ambiguous, and calibrated to the appropriate base or plus level.

The concept and creation of user-oriented proficiency scales, comprised largely of can-do statements, is not altogether new, even in the United States. The ILR had already been adapted to the requirements of the U.S. and its NATO partners in the Standardization Agreement 6001 (2010). In the STANAG 6001, the same base and plus levels (0+ through 5) are described, largely in the form of can-do statements that reflect communicative priorities for military personnel.⁷ Although experienced ILR users might find the STANAG 6001 over-specified, even to the exclusion of proficiency profiles that would otherwise fall within the corresponding ILR levels, it has for decades been used to provide foreign language standards for military personnel of the NATO partner nations.

To avoid the problems of over-specification for proficiency acquisition and assessment, an alternative would be to use descriptor banks as a supplement to the ILR proficiency definitions. Paired with the ILR Skill Level Descriptions, banks of proficiency descriptors—each calibrated to its respective base or plus level—would offer guidance for agencies and individuals regarding the functions most relevant to their professional domain. Rather than an alternative to general proficiency, calibration of the descriptors would make explicit which threshold skills from the ILR Descriptions are needed as a linguistic foundation for more specialized language tasks. As USG agencies as well as business and NGOs show more interest in job-relevant language competence, the Interagency Language Roundtable is optimally situated and experienced to provide leadership in the creation of user-oriented descriptor banks supplementary to the the ILR Descriptions. Now that an ever more diversified global economy and strategic environment has become the leading reason for learning other languages, *general proficiency* is in danger of becoming an antiquated expression—if not exactly a foreign word.

NOTES

1. Council of Europe (2001), *Common European Framework of Reference for Languages: Learning, teaching, assessment*. Cambridge: Cambridge University Press. In this paper, the acronym “ILR” is used to refer to the Interagency Language Roundtable Language Skill Level Descriptions. Whenever I mean the Roundtable as an organization with representatives from its member departments and agencies, I use its full name, the Interagency Language Roundtable.
2. For a more detailed critique of the “real-life” approach to ILR and ACTFL proficiency assessment, see Bachman (1990), pp. 308-312; and Bachman and Sauvignon (1986).
3. Examples of proficiency scales related to CEFR include those of Evaluation & Accreditation of Quality in Language Services (EAQUALS) and the Association of Language Testers in Europe (ALTE).
4. Lowe (1988) traces the need for a fundamentally new language testing system to conditions in university foreign language departments: “Proficiency evaluation was developed when academia focused heavily on literature” (p. 23). The Modern Language Association’s lopsided priorities, heretofore favoring literature and cultural criticism to all but exclusion of language acquisition, are undergoing some correction with the newly organized Ad Hoc Committee on Foreign Languages. Nevertheless, the committee’s inaugural declaration denigrates what it calls the “instrumental” view of language, as “a skill to use for communicating thought and information”, in favor of “a broad, intellectually driven approach to teaching language and culture in higher education.” In other words, the demand for communicative language skills for life and work in foreign language environments is likely to remain a low priority for professors focused on intercultural scholarly discourse. See Ad Hoc Committee on Foreign Languages (2015), “Foreign Languages and Higher Education: New Structures for a Changed World,” *Modern Language Association*, retrieved from <http://www.mla.org/flreport>.
5. Lowe (1988) describes the urgency that forced language-testing units to proceed with test and metric development rather than await conclusive research. After recalling a comment from a testing chief expressing impatience with the slow pace of “theory”, Lowe assures his readers “that the requisite expertise” to create a metric for rating foreign language proficiency “existed within the government, and that it derived from academia” (p. 16). As evidence, he cites the Ivy League credentials of Claudia Wilds, “a primary designer of oral proficiency assessment procedures and later head of testing at the FSI School of Language Studies.”
6. See North and Schneider (1998) concerning the process adopted to gather descriptors, or “criterion statements” from a number of proficiency scales already in existence (pp. 226-228). They then describe the process in which the descriptors were calibrated to a CEFR level, by submitting the aggregate

ratings of individual language teachers to quantitative analysis using the Rasch psychometric model.

7. See Military Committee Joint Standardization Board (2010). It is worth noting that, whereas only a few of the linguistic tasks elicited during the Oral Proficiency Interview, as it is administered at DLIFLC, can be found in the the ILR Descriptions, nearly all may be found in or inferred from the STANAG 6001, which can be explained by the fact that many of the same individuals were involved in both writing the STANAG 6001 document and designing test content and procedures for OPI 2000.

REFERENCES

- Ad Hoc Committee on Foreign Languages (2015). Foreign languages and higher education: New structures for a changed world. *Modern Language Association*. Retrieved from <http://www.mla.org/flreport>
- Alderson, J. C. (1991). Bands and scores. In J.C. Alderson, & B. North (Eds.), *Language testing in the 1990s* (pp. 71–86). London: Modern English Publications/British Council/Macmillan.
- American Council on the Teaching of Foreign Languages (2012a). *ACTFL proficiency guidelines 2012*. White Plains, NY: ACTFL.
- American Council on the Teaching of Foreign Languages (2012b). *Performance descriptors for language learners*. Alexandria, VA: ACTFL.
- Bachman, L. F. (1990). *Fundamental considerations in language testing*. Oxford: Oxford University Press.
- Bachman, L.F., & Savignon, S. J. (1986). The evaluation of communicative language proficiency: A critique of the ACTFL oral interview. *Modern Language Journal*, 70 (4), 380-390.
- Council of Europe (2001). *Common European framework of reference for languages: Learning, teaching, assessment*. Cambridge: Cambridge University Press.
- Interagency Language Roundtable (n.d.). *Interagency Language Roundtable language skill level descriptions – speaking*. Retrieved from <http://www.govtilr.org/Skills/ILRscale2.htm>
- Interagency Language Roundtable (n.d.). Self-assessment*. Retrieved from <http://www.govtilr.org>
- Lowe, P. (1982). *ILR handbook on oral interview testing*. Washington, DC: DLI-ILR Oral Interview Project.
- Lowe, P. (1988). The unassimilated history. In P. Lowe, & C. Stansfield (Eds.), *Second language proficiency assessment: Current issues* (pp. 11-51). Englewood Cliffs, NJ: Prentice Hall.
- Lowe, P., & Stansfield, C. W. (1988). Introduction. *Second language proficiency assessment: Current issues* (pp. 1-10). Englewood Cliffs, NJ: Prentice Hall.
- Military Committee Joint Standardization Board (2010). *Standardization agreement (STANAG) 6001 – Language proficiency levels* (4th ed.). Brussels: NATO Standardization Agency.
- National Council of State Supervisors for Languages, & the American Council on the Teaching of Foreign Languages (2014). *Can-do statements: Progress indicators for language learners*. Alexandria, VA: ACTFL.
- North, B., & Schneider, G. (1998). Scaling descriptors for language proficiency scales. *Language Testing*, 15(3), 217-263.
- Rice, F. A. (1959). The Foreign Service Institute tests language proficiency: State Department examines government personnel. *The Linguistic Reporter* 1 (2) (suppl.1), 4.

Upshur, J. A., & Turner, C. E. (1995). Constructing rating scales for second language tests. *ELT Journal*, 49(1), 3-12.

AUTHOR

Don Holman, Ph.D., Professor / Director of Hurlburt Field Language Training Detachment, Defense Language Institute Foreign Language Center.

Using Writing to Teach Pronunciation: An Experimental Fourth-Year University Spanish Phonetics/Phonology Course

MICHAEL D. HUBERT
Washington State University

DONNY VIGIL
University of St Thomas

Adult second/foreign language acquisition is extremely challenging for many learners, with pronunciation often the one aspect in which otherwise very proficient language users fall short. Although a great deal of research has found that formal instruction in phonetics/phonology may improve learner pronunciation, no study has yet addressed the effects that the addition of formal written assignments to classroom instruction in phonetics/phonology may have on the development of learner pronunciation. This study investigated the effects of explicit instruction involving extensive writing assignments on learner acquisition of the Spanish voiced approximants [βðy] and the glides [jw]. Twenty eight native English-speaking students enrolled in two different advanced seminars on Spanish phonetics/phonology were tested on their ability to produce the targeted sounds at the beginning and end of a university semester. Results indicated that moderate benefits may accrue by using writing to teach pronunciation for some learners in this type of course.

Keywords: *acquisition of pronunciation, Spanish phonemics, writing assignments, instruction of phonetics/phonology*

The observation that language learning becomes more difficult after childhood is well-documented within second language acquisition (SLA) research. Current versions of Lenneberg's original (1967) critical period hypothesis (CPH) (Bley-Vroman, 1989; Dekeyser, 2000; Scovel, 2006; Birdsong, 2006) highlight the highly variable nature of language acquisition that takes place after biological maturity. In short, these claim that language acquisition by adults is still possible, but that (1) SLA is characterized much more by failure than by success (Han, 2004), and (2) levels of second language (L2) proficiency attainment vary greatly from learner to learner (e.g., Scovel, 1988; Singleton, 2012). In the area of L2 pronunciation in particular, CP effects are believed to start early in life, (Oyama, 1976; Scovel, 1988; Flege, 1992a, 1992b, 1995), and to (severely) limit adult learners' ability to achieve native-like levels of second language proficiency (Bongeaerts, Plankton, & Schils, 1997; Han, 2004). Pronunciation is often the one skill in which otherwise very proficient adult L2 users fall short. Fortunately, research conducted with advanced L2 learners has suggested that formal instruction in the phonology of the target language (TL) may improve pronunciation and overcome some CP effects (i.e., Zampini, 1994; González-Bueno, 1995; Elliott, 1997; Díaz-Campos, 2004; Lord, 2005; Gonzalez-Bueno & Quintana-Lara, 2011; Zárates-Sánchez, 2011).

However, no study has yet addressed the positive effects that the addition of writing to formal phonetics/phonology instruction may have on the development of learner pronunciation accuracy. Where *Writing-to-Learn* activities, those whose primary purpose is to teach aspects of language development other than writing itself (Lefkowitz & Hedgcock, 2009; Hedgcock & Lefkowitz, 2011), tend to predominate in the traditional foreign language (FL) classroom (Scott, 1996; O'Donnell, 2007; Hubert, 2011; Reichelt, Lefkowitz, Rinnert, & Schultz, 2012; Hubert, 2014) and are frequently used by teachers to introduce and/or reinforce different elements of the TL. Writing assignments such as these are often used to help improve learner speaking and/or grammar, but do not seem to be utilized to teach or reinforce TL phonetics/phonology. This study investigates the effects of a pedagogical intervention involving both explicit instruction and extensive writing assignments on learner acquisition of the Spanish voiced approximant allophones /bdg/ → [βðɣ]¹ (e.g., *el vaso* [the cup] /el#baso/ → [el.'βa.so]), *los dedos* [the fingers] /los#dedos/ → [loz.'ðe.ðos], *el gato* [the cat] /el#gato/ → [el.'ya.to]) and the glides /iu/ → [jw] (e.g., *buen* 'good' /buen/ → ['bwen], *ciudad* 'city' /siudad/ → [sju.'ðað]). Spanish is, in a general sense, a highly phonetic language, especially when compared to English, as far more sounds have a one-to-one correspondence to a letter in the written alphabet in Spanish than they do in English. Although this fact may make pronunciation training easier at the beginning levels, several important exceptions to this rule cause difficulty for more advanced learners, including the phonological rules governing the use of Spanish [βðɣjw]. Due to the fact that incorrect use of the closest English sound to [βðɣjw] does not generally result in loss of intelligibility, and these sounds have been judged to be of relatively low salience for first language (L1) English

learners of Spanish (Lord, 2010; Gonzalez-Bueno & Quintana-Lara, 2011), these sounds were excellent candidates for an experiment in the consciousness-raising effects of writing on student learning.

REVIEW OF THE LITERATURE

The Spanish Voiced Approximants [βðɣ]

Although the phonemic inventories of both Spanish and English contain the phonemes /bdg/ and their occlusive allophones [bdg], Spanish may be considered to have a spirantization rule that produces a series of approximant allophones in cases where /bdg/ do not appear at the beginning of the breath group or directly following a nasal consonant (also following a lateral sound for /d/ only). In the majority of environments, the student of Spanish needs to acquire [βðɣ] in addition to [bdg] (Hammond, 2001).

After pause, nasal (also lateral for /d/): [bdg]

- | | | |
|-----------|-----------|---------|
| (1) boca | ['bo.ka] | [mouth] |
| (2) datos | ['da.tos] | [data] |
| (3) gato | ['ga.to] | [cat] |

All other contexts: [βðɣ]

- | | | |
|---------------|---------------|-------------|
| (4) la boca | [la.'bo.ka] | [the mouth] |
| (5) los datos | [loz.'ða.tos] | [the data] |
| (6) el gato | [el.'ya.to] | [the cat] |

After nasal:

- | | | |
|------------|------------|-----------|
| (7) ambos | ['am.bos] | [both] |
| (8) andar | ['an.dar] | [to walk] |
| (9) inglés | [in.'gles] | [English] |

After lateral:

- | | | |
|------------|------------|-----------------|
| (10) aldea | [al.'de.a] | [small village] |
|------------|------------|-----------------|

Although English does make use of most of these sounds (the exceptions being [βɣ]), the phonological rules governing their use are very different and, therefore, are not produced in the same environments. L1 English learners often have trouble recognizing and producing these sounds in the correct environments, and instead tend to produce the occlusive allophones in situations in which the approximant is required. Although this error does not generally contribute to a loss of intelligibility, the non-target like use of these allophones can lead to the perception of a foreign accent (Lord, 2010).

Spanish [jw] and Allophonic Glide Formation

The Spanish high vowels /i/ and /u/ undergo glide formation in non-syllable-initial, unstressed environments when in direct contact with another vowel (Hammond, 2001), and result in the production of [j] and [w] in these cases, even across the word boundary. However, the vast majority of FL Spanish students never learn this rule. Instead, they are taught that written *i* and *u* are always pronounced as [i] or [u]. It also seems that many instructors, both native and non-native, are also unaware of this fact, even if they do produce the surface forms correctly. Certain instances of this sound change tend to be unproblematic for FL Spanish learners, such as *bien* [bjén] [good] which is almost never pronounced as *[bi.én]², we believe due to the fact that students hear this word pronounced frequently in class, whereas other less frequent words such as *también* [tam.bjén] [also] are commonly pronounced as *[tam.bi.én]. This is especially true across the word boundary, for phrases such as *mi alma* [mjál.ma] [my soul] pronounced as *[mi.al.ma].

The Effects of Instruction on the Acquisition of Pronunciation

A growing number of studies have suggested that explicit pronunciation instruction can aid in the acquisition of L2 pronunciation (e.g., Bjarkman, 1986; Terrell, 1989; Zampini, 1994; González-Bueno, 1995; Castino, 1996; Elliott, 1995, 1997; Cenoz & Lecumberri, 1999; Derwing & Rossiter, 1998; Archibald, 1998; Major, 1998; Moyer, 1999; Arteaga, 2000; Burgess & Spencer, 2000; Derwing & Rossiter, 2003; Díaz-Campos, 2004; Lord, 2005, 2010; Gonzalez-Bueno & Quintana-Lara, 2011; Zárata-Sández, 2011). These add to the growing body of evidence that formal instruction may make a difference in helping to overcome critical period (CP) effects (Han, 2004).

The Effects of Instruction on Spanish [βðγ]

Several studies have investigated both the perception and production of [βðγ] by native speakers of English. First, Elliott (1997), working with students enrolled in an intermediate Spanish course, found that formal instruction improved learner accuracy in the production of [βðγ]. However, accuracy decreased when students' attention was shifted from an explicit focus on pronunciation to communication. Lord (2005) investigated the effects of explicit instruction and self-analysis on the acquisition of nine Spanish phonemes, including [βðγ], on students enrolled in an advanced Spanish phonetics/phonology course. Analysis of pre- and post-instructional participant pronunciation recordings indicated that explicit instruction and self-analysis led to significant gains in accuracy for all targeted sounds.

Lord (2010) also investigated the role of immersive study abroad and explicit instruction in phonology in the acquisition of [βðγ]. Students taking part in a 10-week summer immersion program, some of them had previously completed an advanced phonetics/phonology seminar, were recorded reading

aloud a passage containing the targeted sounds both before and after this summer program. Her findings indicated that both explicit instruction and study abroad may have a positive effect on the acquisition of these sounds, with some sounds seeing as much as a 40% increase in target-like production.

Gonzalez-Bueno and Quintana-Lara (2011) specifically investigated whether students at different levels of proficiency are aware of the spirantization rule governing the Spanish voiced stops explained in the section of the Spanish voiced approximants [βðɣ]. University students at low, intermediate, and advanced levels of proficiency were tested on their ability to (1) detect the target-like pronunciation of [βðɣ], and (2) pronounce these sounds in a target-like manner. Participants listened to a series of 30 tape-recorded Spanish words, each containing one of the obligatory contexts for the application of this rule (e.g., *dedo* ‘finger’ [de.ðo]). In half of these contexts, the sound was produced as [bdg], and in the other half they were pronounced as [βðɣ]. Participants were asked to judge each word as having been pronounced with spirantization. Results indicated that intermediate proficiency learners had started to recognize [ð] and [ɣ], with some trouble recognizing [β], and that advanced learners recognized [ð] and [ɣ] no differently from the intermediate learners, but were better at recognizing [β]. In regards to production, participants were asked to read an 80-word paragraph containing 38 instances of the targeted sounds, 13 words containing [β], 18 with [ð], and seven with [ɣ]. Results indicated that these participants were able to apply the spirantization rule at the low proficiency level for [ɣ], and at the intermediate level for [β] and [ð].

The Effects of Instruction on Spanish [jw]

There is currently very little research available on Spanish diphthong (and nasal allophone) acquisition. One such study is Lord (2005), who investigated the effects of explicit instruction and self-analysis on the acquisition of [jw] both within and between words, following the rule explained in the section of Spanish [jw] and allophonic glide formation. Participants were students enrolled in an upper-division university Spanish phonetics course. Two voice recordings of each student were made, one at the beginning and the other at the end of the semester. Students received explicit instruction in the phonological rules governing these sounds, and completed self-analyses of their own pronunciation during the semester. Results indicated a statistically significant increase in the accuracy with which learners produced these glides in obligatory contexts.

Zárate-Sánchez (2011) conducted a study with university students enrolled in beginning, intermediate, and advanced Spanish courses. Each participant was asked to complete a written syllabification exercise in which words containing both [i] and [u] and [j] and [w] were found. After pronouncing the word silently to themselves and attempting to sound out the syllable divisions, participants were instructed to draw slashes between written letters to mark these syllable divisions. Results indicated that learners at all three levels of proficiency showed a statistically similar propensity towards producing [i] and

[u] in contexts in which native speakers of Spanish produce [j] and [w], including pronouncing words like *función* [function] as *[fun.si.'on] instead of the target-like [fun.'sjon] and phrases such as *mi alma* as *[mi.'al.ma] instead of ['mjal.ma].

A Model of Acquisition: Using Writing to Foster Pronunciation

Perhaps the most widely-accepted model of L2 pronunciation acquisition was developed by Flege (1992a, 1992b, 1995), who proposed that the processes and/or mechanisms guiding the successful L1 acquisition of new phonetic categories remain intact past the critical period, and that it is perception that determines a learner's ability to produce TL sounds; in short, any TL sound must first be perceived before it can be produced, and that production is guided by the perceptual representations of sounds learners store in their long-term memory. In addition, Flege postulated that the elements comprising both the L1 and TL phonetic systems exist in a *common phonological space* and therefore can mutually influence one another. Where earlier research had argued that the primary constraints on the post-CP acquisition of pronunciation were primarily biological/neurological in nature (e.g., Molfese, 1977; Scovel, 1988; MacNeilage & Davis, 1990; Kent, 1992), and had assumed that adult learners lose the ability developed during childhood to perceive and assign categories to sounds, Flege instead proposed an explanation based on the premise that adult learners not only continue to possess the necessary neuromuscular ability to produce new sounds, but that they are also capable of perceiving new TL sounds and of organizing these sounds into new phonetic categories.

Flege proposed that (1) learners are more likely to form new phonetic categories for TL sounds of greater perceived dissimilarity from the closest L1 sound, (2) category formation becomes increasingly difficult during the CP as representations for similar L1 sounds develop, (3) learners will assimilate, or *merge*, similar L1-TL sounds in cases in which TL categories are not correctly developed, and (4) that given enough time and exposure, learners may come to dissimilate these pairs of L1-TL sounds and establish correct TL categories. In sum, Flege's Speech Model proposes that TL learners can, given enough time and enough of the right kinds of input, correctly perceive the phonetic properties of L2 speech sounds and store this information in their long-term memory, despite their initial incorrect perceptions of different TL sounds.

The noticing hypothesis (Schmidt 1990, 2001) argues that input cannot enter into short-term memory (i.e. be converted to *intake*) unless it is noticed at a conscious level. Most empirical studies on the topic have upheld the noticing hypothesis (Schmidt, 2010), and a great deal of evidence remains to suggest that conscious attention to form is necessary, at least on some level, for the acquisition of most language elements to occur for the majority of learners. We would further posit that not only do learners need to be able to unconsciously or semi-consciously perceive and categorize new TL sounds, but that many will also need to, and would greatly benefit from, pronunciation instruction that encourages a completely conscious knowledge of TL phonetic categories.

Where speaking practice might seem the most obvious choice for practicing newly-presented concepts, we believe that learner writing in the TL, in conjunction with explicit instruction in phonetics/phonology, may foster the acquisition of TL pronunciation even more so than practice via speaking alone.

Writing in the TL affords the learner several advantages over speaking that may promote noticing and/or metalinguistic reflection. First, writing may enhance a learner's opportunity to notice different TL elements that would elude them during speech. The learner is simultaneously responsible for both TL production and processing during the speech act, which generally takes place at a much higher rate than does the act of writing. The resulting cognitive demands placed on the learner's attentional resources are much higher than those required during writing. An important group of SLA researchers (i.e., Krashen, 1985, 1989; Schmidt, 1990; Robinson, 1995; VanPatten, 2002) has described the human language acquisition device as a limited-capacity system in which the L2 learner, particularly at beginning levels, is unable to attend to all non-automatized (Dekeyser, 2001) TL elements at the same time. The mind is put under a great deal of strain when required to produce the TL under limited time constraints.

As speech is a phenomenon that happens *on-line* or *in real-time*, successful speech must be planned, formulated, and articulated with considerable speed. These demands on the learner's developing interlanguage may make the perception and/or noticing of non-meaning-bearing TL elements next to impossible (VanPatten, 2002). Even if learners are made explicitly aware of the way in which they *should* be producing certain TL phonetic categories, they may find themselves unable to do so, or perhaps acquiring this ability at a slower rate, due to the demands placed on the language system during the speech act.

Writing, on the other hand, gives the learner more time to think about the topic at hand, and more time to plan, edit, and revise utterances. Much more so than speaking, writing allows learners to focus their explicit attention on targeted aspects of the TL, which can very possibly include a focus on phonetic categories and TL sounds, and to do so for as long as needed. We believe that these hypotheses, up to this point applied only to grammar and vocabulary, are also applicable to the acquisition of pronunciation, as TL writers have more time to reflect on their own language use, more time to plan their utterances, more time to brainstorm ideas, edit, and revise their output as they produce a written text. In addition, we believe that the act of producing formal writing *about* pronunciation may lead to additional metalinguistic reflection due to the increased level of formality and level of detail required in the production of a formal written text on TL phonetics/phonology, as FL students are obliged to revisit their memories and understanding of explicit explanations of material.

Additionally, writing may simply provide learners additional opportunity to produce the TL that they simply do not have via speaking alone. Many FL learners have only very limited access to authentic native speaker input, and even less opportunity to produce TL output. Writing-to-learn assignments, whose primary purpose is to teach aspects of language

development other than writing itself, including traditional written homework designed as additional language practice, journal writing, and/or formal essays, take student involvement with the TL beyond the limited walls of the classroom by greatly extending the time these learners must spend to successfully complete their FL studies. Particularly in the case of formal essays and other assignments requiring learner editing and revision, a great deal of additional time outside of class may be required to complete these assignments, arguably time that learners would not otherwise spend engaged in TL use.

METHODOLOGY

Research Questions

The following research questions guide the present study:

1. Can an advanced phonetics/phonology course including extensive writing assignments aid native English-speaking students in their pronunciation of Spanish [β], [ð], [γ], [j] and [w]?
2. Will this type of course design lead to greater gains in pronunciation accuracy than a similar course design that does not include extensive writing assignments?

Participants

Participants in the present study were taken from two different 400-level seminars on Spanish phonology offered at two different universities, one in the northwestern United States (serving as the experimental group, $n=19$), and the other in the southwest U.S (serving as the control group, $n=9$). All participants whose data are included in this study were native speakers of American English who grew up in the United States. However, both classes did enroll additional students who also completed all study elements. These included two L1 Spanish-speaking learners in the experimental group, and five L1 Spanish-speaking learners, two simultaneous bilingual learners, and one L1 Brazilian Portuguese speaker from the control group. Data from these non-L1 English students are not included in the present study.

All experimental participants had completed the equivalent of at least three years of traditional university Spanish language instruction in which no formal pronunciation instruction, including the science of acoustic and articulatory phonetics, the phonetics/phonology distinction, allophonic variation, phonetic transcription, and the concept of the sonority hierarchy, was offered. Overall communicative proficiency levels for participants in both experimental and control groups were fairly high, but not assessed nor recorded as part of the present study. It was assumed that, by virtue of their enrollment in these two 400-level seminars, all participants possessed a minimum level of Spanish proficiency. The present course represented for many of these students the last university Spanish language course they would take during their university careers.

Course Description – Control Group

This course met for a total of 50 minutes three times per week over the course of a 16-week university semester and was taught by one of the authors of this article. The class format consisted of a mixture of direct lecture in the TL, weekly readings in the learners' native language (English) from the course textbook, *The Sounds of Spanish: Analysis and Application* (Hammond, 2001), individual and group in-class phonetic transcription practice, whole-class discussions of course material in the TL, four formal in-class exams in which students were directly assessed on their declarative knowledge of course material, but not on their ability to use the targeted sounds, a dialect analysis and identification assessment in which students listened to native Spanish speakers from different parts of the world read from a text containing examples of the entire phonemic inventory of Spanish, and a final class project. This project consisted of two parts: (1) the recording and transcription of a native speaker of Spanish completing a reading passage containing instances of the entire phonemic inventory of Spanish; and (2) the recording and transcription of themselves reading the same passage.

Course lectures first introduced students to the science of acoustic and articulatory phonetics, the phonetics/phonology distinction, allophonic variation, phonetic transcription, as well as the concept of the sonority hierarchy. After this initial training, a formal phonological analysis of the phonemic and allophonic inventories of the Spanish language were presented, including a complete articulatory description of each vowel and consonant, as well as their distribution, orthographic representations, and allophonic variants. Each consonant and vowel description ended with a comparison between Spanish and American English. The course then presented a description of the suprasegmental features of the Spanish language, and finished with an overview of the major dialectal areas of the Spanish-speaking world and their unique phonological features, making use of the book *Spanish Pronunciation in the Americas* (Canfield, 1981), also written in English.

Course Description – Experimental Group

This course met 75 minutes two times per week during a 16-week university semester, and was taught by the other author of this article. The same two textbooks read by control group students were also used in this course. Therefore, the content of the lectures were essentially the same across the two study groups, with no significant variations worthy of mention here.

This course differed from the curriculum used with the control group in that much more formal, multi-draft writing was completed by students in place of other types of assignments, and that these written assignments comprised 55% of the overall course grade. First, each week students were required to complete a 1-page formal summary of the assigned reading for that week in Spanish written in the TL. These summary assignments were designed to give

students the additional consciousness-raising opportunities introduced in the literature review. As they were due at the end of each week, and could be based on their understanding of the different issues gleaned from both course readings and in-class lectures, these assignments were designed to give learners the opportunity to revisit, organize, and solidify their understanding of the material covered that week. These summaries were graded primarily on the success with which students correctly represented the material presented in the different readings. Those that did not adequately represent the targeted information in their summaries were required to submit a second draft for a higher grade. Students received frequent feedback on their writings throughout the course, most of which focused directly on their understanding of the material instead of on language form or writing convention.

Students in this group also completed a multi-stage final research project in which they examined the sounds produced by native Spanish speakers from selected dialects (students selected dialects to study that interested them) and compared their own performance to that native speaker standard. This project required an annotated bibliography and working outline turned in for feedback by the 8th week of instruction, an initial draft turned in to the instructor for feedback, an oral presentation of their research to the rest of the class, and a final draft that brought together information and feedback received from all previous elements of the assignment. This assignment was similarly designed to provide learners with multiple consciousness-raising opportunities to notice those Spanish sounds that they themselves may not have been producing correctly at the time.

Lastly, the formal memorization exams offered in both courses were conducted differently. Where four in-class exams were given in the control group, two take-home essay exams were used in the experimental group. These exams provided learners with more opportunities to organize, solidify, and showcase their understanding of course material, and to take as much time to conduct as much research as was necessary to produce this written document.

Data Collection

The collection of pronunciation data was conducted in two stages that represent, in short, pre- and post-course treatment measures of participant Spanish-language pronunciation proficiency. The first collection took place during the first week of the semester and the second during the last week of instruction, before final exams were administered the following week. All recordings were made using a high-quality digital voice recorder which creates 320Kbps digital sound files. These files were transferred from the recorder to a secure server where they were later accessed for analysis.

During the first stage of data collection, a recording was made of each participant, in which s/he read aloud a 300-word passage specifically designed to elicit the pronunciation of the Spanish sounds targeted by the present investigation. See the appendix to this article for a copy of this reading passage. Each participant made an out-of-class, one-on-one appointment with course

instructors to make the recording in a quiet, private environment. Participants were given the previously unseen reading passage and asked to immediately begin to read without any advance preparation. The course instructors remained present during the recording. No repeat recordings were allowed, and each participant was informed of this prior to receiving the reading passage. These recordings were later transcribed for analysis, targeting their use of [βðγ] and [jw].

The second stage of data collection took place after all course components had been completed by students, with the exception of the final draft of the research project conducted by the experimental group, and consisted of a series of recordings in which participants again read the same 300-word passage. Although the use of an identical reading passage for the purposes of showing learner gains may be frowned upon by some, the argument is put forth that the *memorization effect* was not a confounding factor in this case, as participants did not have access to the passage at any point during the 15 weeks which passed between pre- and post- recordings. These recordings were also transcribed for analysis, again targeting the use of [βðγ] and [jw].

Data Analysis

Each participant's pre- and post-instructional recording was analyzed for the target-like pronunciation of each of the targeted Spanish sounds found in the data collection reading. Analysis was carried out in terms of a binary target-like/non-target-like distinction. These judgments were made by two raters on the basis of place and/or manner of articulation of each sound. At the time, both raters were enrolled as graduate students studying Spanish literature, and both had only a basic level of understanding of Spanish phonology. One was a native English speaker from the United States, and the other was a native Spanish speaker from Central America. These raters both received training from one of the authors of this article on the distinction between the sounds under investigation in this study, including a series of pilot ratings prior to their rating of the present samples.

After all data collection had been completed, digital recordings from the four study groups (experimental pre-test, control pre-test, experimental post-test, and control post-test) were combined in a single computer folder on a password-protected server space for rating. The raters were given access to the recordings for the limited time it took them to complete their ratings. Due to the fact that all recordings had previously been assigned a random six-digit number, and to the fact that all recordings were delivered for analysis at the same time, the raters did not know from which group (experimental or control), nor from which test (pre- or post-treatment), they were rating at any given time. A master list containing a list of numbers associated with each of the four recording groups was kept confidential and not shared at any time with the raters, and was afterwards used to assign recording scores to each of the four groups.

Raters listened to the recordings, noting the number of instances of target-like and non-target-like production of each of the targeted sounds, one

sound at a time. All empirical data for this study come from these ratings. Global pre- and post-instructional accuracy rates and standard deviations for both experimental and control groups were calculated for each sound. In the case of rater disagreement, the results were averaged.

RESULTS

Inter-rater Reliability

Inter-rater reliability (IRR) correlations are as follows: overall IRR was calculated at 0.897, with [βδγ] IRR at 0.925, [j] IRR at 0.915, and [w] IRR at 0.852.

Pre- and Post-treatment Accuracy Rates

Results of the pre- and post-treatment ratings of participant production of the targeted sounds are presented in Table 1.

Table 1
Production Accuracy by Sound Type

	<i>Experimental Group (n=19)</i>		<i>Control Group (n=9)</i>	
	<i>Pre-Test</i>	<i>Post-Test</i>	<i>Pre-Test</i>	<i>Post-Test</i>
<i>[β] (n= 18)</i>	Mean 72.22% σ - 0.177	Mean 74.56% σ - 0.166	Mean 68.83% σ - 0.124	Mean 67.28% σ - 0.205
<i>[δ] (n= 70)</i>	Mean 83.12% σ - 0.160	Mean 85.83% σ - 0.117	Mean 91.75% σ - 0.099	Mean 91.90% σ - 0.099
<i>[γ] (n= 7)</i>	Mean 72.56% σ - 0.286	Mean 79.32% σ - 0.260	Mean 79.37% σ - 0.261	Mean 83.33% σ - 0.200
<i>[βδγ] total</i>	Mean 75.97% σ - 0.163	Mean 79.90% σ - 0.133	Mean 79.98% σ - 0.147	Mean 80.84% σ - 0.148
<i>[j] (n= 25)</i>	Mean 71.79% σ - 0.160	Mean 80.11% σ - 0.128	Mean 74.44% σ - 0.247	Mean 78.44% σ - 0.198
<i>[w] (n= 8)</i>	Mean 43.09% σ - 0.299	Mean 53.95% σ - 0.246	Mean 44.44% σ - 0.292	Mean 59.03% σ - 0.203

In order to determine if the difference between pre- and post-test scores was statistically significant, a one-way ANCOVA was calculated for each group of sound scores. These data are summarized in Table 2. Calculation of the analysis of covariance between the experimental and control group pre- and post-tests revealed no statistically significant change in the accuracy of these groups' production of any of the targeted sounds that can be attributable to the present treatment. Whereas 69.5% of the variability observed in post-test

accuracy rates for [βðʏ] were found to be accounted for in the pre-test scores, this number dropped to 33% for [j] and only 7.5% for [w]. However, p values for all sounds were found to be too large to be considered statistically significant, although improvement for [w] came rather close at $p=0.075$.

Table 2

One-way ANCOVA Results by Sound Type

Experimental Group (n=19), Control Group (n=9), df=(1,26)

<i>Sound</i>	<i>F</i>	<i>p</i>	<i>Partial Eta Squared</i>
[β, ð, ʏ]	0.392	0.537	0.695
[j]	0.307	0.585	0.33
[w]	1.475	0.236	0.075

DISCUSSION AND CONCLUSIONS

Research Question 1

The answer to our first research question is complicated and requires a nuanced analysis in order to fully appreciate what the present data can show us. First, although the accuracy with which the experimental group produced each of the five targeted sounds did improve in each case, this improvement was not statistically significant at the group level. This might suggest that the treatment was ineffective in helping students to improve their pronunciation. However, it is of vital importance here to further highlight the highly variable nature of our results. Where the majority of participants showed only little improvement (if any) in the accuracy of their production of the five sounds, a small number of participants in the experimental group showed much more dramatic and substantial gains, particularly for the glides [jw], as illustrated in Figures 1 and 2.

strongly support the CPH, in that they provide additional evidence that learning beyond the critical period for language acquisition is still possible, but that it is highly variable [across learner populations] and less successful than learning which takes place within the critical period (Han, 2004).

Research Question 2

With regard to our second research question, whether or not a pronunciation course involving extensive writing assignments would lead to greater gains in pronunciation accuracy when directly compared to a similar course without such assignments, the present data found no statistically significant difference between the two groups. However, we did find some individual differences that may suggest that pronunciation instruction involving writing assignments is perhaps more effective at reaching a small number of more receptive students. As was the case with the experimental group participants, a small number of learners in the control group showed much greater gains than the rest. For [βðγ], the most-improved control group participant showed a gain in accuracy of 8%, with four others showing gains of between 2%-5%. For [j], five of the nine participants showed slight improvement of between 2%-6%, with one participant showing significant improvement of 14%. For [w], five showed improvements of 25%, 31.5%, 43.5%, 62.5%, and 63%. Among experimental group participants, the most-improved participant for [βðγ] showed an improvement of 36%, with three others showing improvements of 11%, 12%, and 13%, respectively. For [j], five participants saw improvements in accuracy at 11%, 18%, 32%, 32%, and 44%. For [w], seven saw dramatic improvements in accuracy at 31%, 31%, 37%, 50%, 56%, 68%, and 81% improvement. In all cases, these smaller numbers of experimental group learners experienced substantially greater accuracy gains than their control group counterparts. However, we must also point out that an even smaller number of participants in both groups experienced a decrease in accuracy between the pre- and post-treatment measures.

Limitations and Future Research

The present study suffers from several important limitations that must be taken into account when gauging the impact of its findings. First, as has traditionally been the case in similar studies of this type, finding a large number of student volunteers to serve as study participants was a challenge. Our total sample size stands at only 28 participants, with 19 in the experimental group and nine in the control group. There were simply not many advanced-level Spanish students interested in taking this type of course available for recruitment at either university. Therefore, the statistical power of these findings is limited, as L2 pronunciation research in general currently suffers from a marked lack of participants. If possible, future research in this area should involve larger groups of students as study participants.

Second, in terms of data collection, a reading used to elicit the targeted sounds from participants did not contain the same number of occurrences of each of the targeted sounds. Two of the sounds, [y] and [w] were elicited a fairly small number of times ($n_y = 7$, $n_w = 14$), further reducing the statistical significance of the generated data. However, due to the fact that these sounds do not share the same rate of occurrence within native speech, this weakness in the reading was ignored in favor of using a text that was both interesting and distracting to participants as they read and produced the targeted sounds.

Lastly, the analysis of the present glide data may be considered incomplete, as raters were not instructed to differentiate between glide formation within words and across the word boundary. This is an important distinction that should be addressed in future research. Anecdotally, we have observed that students seem to have more trouble applying these rules across the word boundary than within the word, possibly due to the affected nature of the speech to which they are exposed in their university classes, or perhaps to the fact that English does not have the same resyllabification rule as does Spanish. Future research into L2 Spanish glide acquisition should study the disparity (should one exist) that exists between the production of glides within words and across the word boundary.

Conclusions

Advanced-level university phonetics/phonology courses have been shown to help students to improve their TL pronunciation (e.g., Elliott, 1997; Lord, 2005), and the general consensus appears to be that explicit pronunciation instruction can aid in the acquisition of L2 pronunciation. Although the present data did not show statistically-significant gains in the accuracy of participants' pronunciation of [βðɣjw] for either the control or experimental groups, a small number of participants in both groups showed gains in accuracy of more than 20% for at least one of these sounds. Whereas 10 of the 19 experimental group participants improved by more than 20% in the accuracy of their pronunciation of [βðɣ], [j] and/or [w], the five control group participants who saw gains of this size all improved only in their pronunciation of [w]. At most, these results suggest that writing may be promoting noticing and/or metalinguistic reflection of more sounds and/or providing a greater percentage of students with additional opportunities to engage with the course material that they do not have in class. In following with Flege's Speech Model, we believe that it is perception that determines a learner's ability to produce TL sounds, that all TL sounds must first be perceived before they can be produced, and that production is guided by the perceptual representations of sounds that are stored in the learner's long-term memory. Although any form of explicit training in phonetics/phonology is likely to help raise learner consciousness of these sounds and therefore help them to perceive these sounds, pronunciation instruction involving extensive writing assignments may be even more useful for those learners that have a greater affinity for visual learning, or for learning through writing. We have shown that this type of instruction is at least as effective as similar instruction

that focuses more on output and overt pronunciation practice, both of which appear to help at least some students to rather dramatically improve at least some aspects of their pronunciation. Hopefully, these findings will help second language instructors to not shy away from having their students write about their pronunciation, or from bringing pronunciation topics into advanced-level writing courses.

NOTES

1. Throughout this article, base phonemes are presented between slashes (e.g. /bdg/, /iu/), surface allophones are presented between brackets (e.g. [βðŷ], [jw]), and written letters that represent these sounds are presented in italics.
2. Non-target-like pronunciation is indicated with an asterisk before the transcription.
3. Although this reading contains a total of 29 instances of /i/ → [j] and 14 instances of /u/ → [w], the following words are not included in the present analysis in order to avoid artificially inflating these results. This is due to the fact that these are fairly high-frequency lexical items that are either (1) presented to students many times with target-like pronunciation and/or (2) are very similar to English cognates:

/i/ → [j] *nov*ec*ientos, cual*qu*ier, qui*en*es, ti*en*en*

/u/ → [w] *frec*u*encia, despu*és*, cual*qu*ier, p*ue*s, mu*er*to, cu*an*tas*

REFERENCES

- Archibald, J. (1998). Second language phonology, phonetics, and typology. *Studies in Second Language Acquisition*, 20(1), 189-211.
- Arteaga, D. (2000). Articulatory phonetics in the first year Spanish classroom. *Modern Language Journal*, 84(3), 339-354.
- Birdsong, D. (2006). Age and second language acquisition and processing: A selective overview. *Language Learning*, 56, 9-49.
- Bley-Vroman, R. (1989). What is the logical problem of foreign language learning? In S.M. Gass & J. Schachter (Eds.), *Linguistic perspectives on second language acquisition* (pp. 41-68). Cambridge: Cambridge University Press.
- Bongearns, T., Plankton B., & Schils E. (1997). Can late learners attain a native accent in a foreign language? A test of the critical period hypothesis. In D. Singleton & Z. Lengyel (Eds.), *The age factor in second language acquisition* (pp. 30-50). Clevedon, England: Multilingual Matters.
- Bjarkman, P. (1986). Natural phonology and strategies for teaching English/Spanish pronunciation. In P.C. Bjarkman (Ed.), *The real-world linguist: Linguistic applications in the 1980s* (pp. 7-115). New York, NY: Ablex Publishing Corporation.
- Burgess, J., & Spencer, S. (2000). Phonology and pronunciation in integrated language teaching and teacher education. *System*, 28, 191-215.
- Canfield, D. (1981). *Spanish pronunciation in the Americas*. Chicago, IL: University of Chicago Press.
- Castino, J. (1996). Impact of phonetics course on FL learners' acquisition of Spanish phonology. *Selecta: Journal of the Pacific Northwest Council on Foreign Languages*, 17, 55-58.
- Cenoz, J., & Lecumberri, L. G. (1999). The effects of training on the discrimination of English vowels. *International Review of Applied Linguistics*, 37, 261-275.
- Dekeyser, R. (2000). The robustness of critical period effects in second language acquisition. *Studies in Second Language Acquisition*, 22, 499-533.
- DeKeyser, R. (2001). Automaticity and automatization. In P. Robinson (Ed.), *Cognition and second language instruction* (pp. 125-151). Cambridge: Cambridge University Press.
- Dekeyser, R. (2003). Implicit and explicit learning. In C. Doughty & M. Long (Eds.), *Handbook of second language acquisition* (pp. 313-348). Boston, MA: Blackwell.
- Derwing, T., Munro, M., & Wiebe, G. (1998). Evidence in favor of a broad framework for pronunciation instruction. *Language Learning*, 48, 393-410.
- Derwing, T., & Rossiter, M. (2003). The effects of pronunciation instruction on the accuracy, fluency and complexity of L2 accented speech. *Applied Language Learning*, 13, 1-18.

- Díaz-Campos, M. (2004). Context of learning in the acquisition of Spanish second language phonology. *Studies in Second Language Acquisition*, 26, 249-273.
- Elliott, A. R. (1995). Field independence/dependence, hemispheric specialization, and attitude in relation to pronunciation accuracy in Spanish as a foreign language. *Modern Language Journal*, 79, 356-371.
- Elliott, A. R. (1997). On the teaching and acquisition of pronunciation with a communicative approach. *Hispania*, 80, 95-108.
- Flege, J. (1992a). The intelligibility of English vowels spoken by British and Dutch talkers. In R. Kent (Ed.), *Intelligibility in speech disorders* (pp. 157-232). Amsterdam: Benjamins.
- Flege, J. (1992b). Speech learning in a second language. In C. Ferguson, L. Menn & C. Stoel-Gammon (Eds.), *Phonological development: Models, research, implications* (pp. 565-604). Timonium, MD: York Press.
- Flege, J. (1995). Second language speech learning. Theory, findings, and problems, In W. Strange (Ed.), *Speech perception and linguistic experience* (pp. 233-277). Timonium, MD: York Press.
- González-Bueno, M. (1995). The effects of formal instruction on the acquisition of Spanish stop consonants. In W. Glass & A.T. Pérez-Leroux (Eds.), *Contemporary perspectives on the acquisition of Spanish, Vol 2: Production, processing, and comprehension* (pp. 57-75). Somerville, MA: Cascadilla Press.
- Gonzalez-Bueno, M., & Quintana-Lara, M. (2011). Students' awareness of Spanish spirantization allophonic rule. In J. Levis & K. LeVelle (Eds.), *Proceedings of the 2nd pronunciation in second language learning and teaching conference*, Sept. 2010 (pp. 180-198). Ames, IA: Iowa State University.
- Hammond, R. (2001). *The sounds of Spanish: Analysis and application*. Somerville, MA: Cascadilla Press.
- Han, Z. (2004). *Fossilization in adult second language acquisition*. Tonawanda, NY: Multilingual Matters.
- Hedgcock, J., & Lefkowitz, N. (2011). Exploring the learning potential of writing development in heritage language education. In R. Manchón (Ed.), *Learning to write and writing to learn in an additional language* (pp. 209-233). Amsterdam: John Benjamins.
- Hubert, M. (2011). The speaking-writing connection: Integrating dialogue into a foreign language writing course. *Electronic Journal of Foreign Language Teaching*, 8(2), 170-183.
- Hubert, M. (2014). Evaluating the role of writing in first and second year university foreign language curriculum. *Journal of Education and Training Studies*, 2(2), 76-86.
- Kent, R. (1992). The biology of phonological development. In C.A. Ferguson, L. Menn, & C. Stoel-Gammon (Eds.). *Phonological development: Models, research, implications* (pp. 65-90). Parkton, MD: York Press.

- Krashen, S. (1985). *The input hypothesis: Issues and implications*. London: Longman.
- Krashen, S. (1989). We acquire vocabulary and spelling by reading: Additional evidence for the input hypothesis. *Modern Language Journal*, 73, 440–464.
- Lefkowitz, N., & Hedgcock, J. (2011). Handle with care: Understanding the fragile linguistic identities of HL and FL learners. Paper presented at the Annual Meeting of the American Association for Applied Linguistics, Chicago.
- Lenneberg, E. (1967). *Biological foundations of language*. New York, NY: Wiley.
- Lord, G. (2005). (How) can we teach foreign language pronunciation? The effects of a phonetics class on second language pronunciation. *Hispania*, 88(3), 557-567.
- Lord, G. (2010). The combined effects of instruction and immersion on second language pronunciation. *Foreign Language Annals*, 43(4), 488-503.
- MacNeilage, P.F., & Davis, B.L. (1990). Acquisition of speech production: Achievement of segmental independence. In W.I. Hardcastle & A. Marchal (Eds.) *Speech production and speech modeling* (pp. 55-68). Dordrecht: Kluwer.
- Major, R. (1998). Interlanguage phonetics and phonology: An introduction. *Studies in Second Language Acquisition*, 20(1), 131-137.
- Molfese, D. (1977). Infant cerebral asymmetry. In S. Segalowitz and F. Gruber (Eds.) *Language development and neurological theory*. New York: Academic Press.
- Moyer, A. (1999). Ultimate attainment in L2 phonology: The critical factors of age, motivation and instruction. *Studies in Second Language Acquisition*, 21(1), 81-108.
- O'Donnell, M. (2007). Policies and practices in foreign language writing at the college level: Survey results and implications. *Foreign Language Annals*, 40(4), 650-671.
- Oyama, S. (1976). A sensitive period for the acquisition of a nonnative phonological system. *Journal of Psycholinguistic Research*, 5, 261-285.
- Reichelt, M., Lefkowitz, N., Rinnert, C., & Schultz, J.M. (2012). Key issues in foreign language writing. *Foreign Language Annals*, 46(1), 22-41.
- Robinson, P. (1995). Attention, memory, and the 'noticing hypothesis'. *Language Learning* 45, 283-331.
- Schmidt, R. (1990). The role of consciousness in second language learning. *Applied Linguistics*, 11, 129-158.
- Schmidt, R. (2001). Attention. In P. Robinson (Ed.), *Cognition and second language instruction* (pp. 3-32). Cambridge: Cambridge University Press.

- Schmidt, R. (2010). Attention, awareness, and individual differences in language learning. In W. M. Chan, S. Chi, K. N. Cin, J. Istanto, M. Nagami, J. W. Sew, T. Suthiwan, & I. Walker (Eds.), *Proceedings of CLaSIC 2010*, Singapore, December 2-4 (pp. 721-737). Singapore: National University of Singapore, Centre for Language Studies.
- Scott, V. (1996). *Rethinking foreign language writing*. Boston, MA: Heinle and Heinle.
- Scovel, T. (1988). *A time to speak. A psycholinguistic inquiry into the critical period for human speech*. Rowley, MA: Newbury House.
- Scovel, T. (2006). Age, acquisition, and accent. In C. Abello-Contesse, R. Chacon-Beltran, M.D. Lopez-Jimenez, & M.M. Torreblanca-Lopez (Eds.), *Age in L2 acquisition and teaching* (pp. 31-48). Bern: Peter Lang.
- Singelton, D. (2012). Ultimate attainment and the critical period hypothesis: Some thorny issues. In M. Watorek, S. Benazzo, & M. Hickman (Eds.), *Comparative perspectives on language acquisition: A tribute to Clive Perdue* (pp. 188-203). Tonawanda, NY: Multilingual Matters.
- Terrell, T. (1989). Teaching Spanish pronunciation in a communicative approach. In P.C. Bjarkman & R.M. Hammond (Eds.), *American Spanish pronunciation: Theoretical and applied perspectives* (pp. 196-214), Washington DC: Georgetown University Press.
- Vanpatten, B. (2002). Processing Instruction: An update. *Language Learning*, 524, 755-803.
- Zampini, M. (1994). The role of native language transfer and task formality in the acquisition of Spanish spirantization. *Hispania*, 77, 470-81.
- Zárate-Sánchez, G. (2011). Speakers' intuitions about L2 syllable structure: Diphthong vs. hiatus resolution in Spanish by English-speaking learners. In L. Plonsky & M. Schierloh (Eds.), *Selected proceedings of the 2009 second language research forum* (pp. 164-181). Somerville, MA: Cascadilla Proceedings Project.

APPENDIX

DATA COLLECTION READING

No, el canibalismo no ha desaparecido...

Hay en la ciudad de México un barrio en el que la gente ha perdido el gusto por comer tamales. Se trata del barrio Portales, en una de cuyas esquinas, frente a una panadería, se instalaba todas las noches una tamalera llamada María Trinidad Ramírez. En julio de mil novecientos³ setenta y uno, después de que su amasio, Pablo Díaz Ramírez, le propinó una más de las brutales palizas con que sin razón acostumbraba atormentarla a ella y a sus hijos, María Trinidad decidió matarlo.

Así lo hizo una noche que Pablo dormía. Fríamente la mujer descuartizó el cadáver e hirvió la cabeza en la lata alcoholera en que acostumbraba cocer sus tamales. Terminada la macabra tarea, marchó a la esquina habitual donde siguió vendiendo su mercancía durante varias noches. Sus clientes formaban legión, y todavía a estas fechas se atormentan pensando en la posibilidad de que la mujer los haya convertido en canibales involuntarios.

En el estado de Veracruz las autoridades intervienen con cierta frecuencia en casos de otro tipo de canibalismo. Algunos totonacas de la región, quienes se matan a machetazos por cualquier dificultad, tienen la costumbre de extraer el corazón a sus víctimas para comerlo en tacos debidamente aderezados con chiles, pues tienen la creencia de que así el espíritu del muerto no regresará a castigarlos.

La inmensa mayoría de los seres humanos se rebela ante la sola idea de poder disfrutar de un banquete a base de uno de sus semejantes. Pero hay una minoría que no le hace el asco a este tipo de guisados. Y no se trata solamente de unos seres anormales de la ciudad de México, de algunos individuos semisalvajes del Veracruz o de unas cuantas tribus aisladas del corazón de África o de Oceanía. Los aficionados a la carne humana pertenecen a todas las razas; están entre nosotros...

AUTHORS

Michael D. Hubert, Ph.D., Associate Professor of Spanish/Graduate Program
Director, Foreign Languages and Cultures, Washington State University.

Donny Vigil, Ph.D., Assistant Professor of Spanish, Modern and Classical
Languages, College of Arts and Sciences, University of St. Thomas.

The Order of Explicit Information in Processing Instruction

ERIN MCNULTY DIAZ

Dickinson College

The two main components of Processing Instruction (PI) are Explicit Information (EI) and Structured Input (SI). Most researchers have concluded that the SI is more responsible for learner gains than the EI (Benati, 2004a, 2004b; VanPatten & Oikennon, 1996; Wong, 2004). However, some researchers have found that EI does significantly impact learner gains (Cox & Sanz, 2015; Farley, 2004b; Farley & McCollam, 2004; Fernández, 2008; Henry, Culman, & VanPatten, 2009). The conventional order of PI is EI followed by SI. The present study seeks to ascertain if the order of EI and SI affects learner gains as measured by interpretation and production scores on four assessments tasks. Two instructional interventions were designed to respond to this research question. The first treatment group received EI and then SI (traditional order) and the second group practiced with SI before encountering the EI.

The grammatical target of the subjunctive/indicative contrast after the temporal adverb cuando provided a number of acquisitional challenges that were addressed in the EI with five input processing principles. A pretest/posttest/delayed posttest (after five months) design was used to measure learner gains. Repeated measures ANOVAs were used to further analyze the scores from these tasks by total, subjunctive, and indicative items. Findings indicated that the order of EI was not statistically significant, but points to explicit feedback as a contributive factor.

Keywords: *Explicit Information, pedagogy, Processing Instruction, Spanish, subjunctive*

INTRODUCTION

Processing Instruction (PI) operationalizes VanPatten's Input Processing (IP) principles in an approach to grammar instruction. The two main components of Processing Instruction (PI), Explicit Information (EI) and Structured Input (SI), have been investigated to determine the causative component of PI. The role of EI is disputed in the literature. Some researchers have found that SI is the component of PI that provokes learner gains (Benati, 2004a, 2004b; VanPatten & Oikennon, 1996; Wong, 2004) and others have demonstrated the EI plays a positive role in learner performance (Cox & Sanz, 2015; Farley, 2004b; Farley & McCollam, 2004; Fernández, 2008; Henry, Culman & VanPatten, 2009). Previous investigations have not studied whether the order of presentation of instructional treatments affects learner gains: EI + SI versus SI + EI.

To this end, two instructional interventions were designed to ascertain if the order of the EI affected learner gains. The targeted grammatical feature for this PI study is the subjunctive/indicative contrast after *cuando* [when]. One treatment group received explicit information then structured input. A second treatment group received structured input then explicit information. The instructional interventions were delivered via a website designed by the researcher. Participants ($n = 14$) were enrolled in an intensive first semester university course without previous exposure to the subjunctive. A pretest/posttest/delayed posttest (after five months) design was used to measure learner gains on production and interpretation tasks. Repeated measures ANOVAs were used to further analyze the scores from these tasks by total, subjunctive and indicative items. Results suggest that the component ordering of PI may contribute positively to learner gains.

The Targeted Grammatical Feature

The targeted grammatical feature for this study is the temporal adverb, *cuando*, plus subjunctive/indicative. This grammatical feature has been previously investigated by a number of PI researchers. Lee and McNulty (2013) focused on the effects of participant language background on acquisition after a PI intervention. Both Henshaw (2011) and McNulty (2012) investigated the two activity types within structured input: referential and affective activities. Discourse-level composition as a result of a PI intervention on the subjunctive/indicative contrast after *cuando* was investigated by Benati, Lee, and McNulty (2010). What is interesting about this grammatical feature in particular is the fact that Spanish allows for both the indicative and the subjunctive after *cuando* depending on the context. The subjunctive is used when the context has not yet happened or is anticipated, as well as when the speaker is not sure when or even if something will happen. When a sentence contains a future reference, it is by its very nature uncertain and the subjunctive encodes the uncertainty of the event as in example (1). Jorge will rest when he

finishes his work at some point in the future; the event of finishing his work has not yet happened or may never happen.

- (1) *Jorge va a descansar cuando termine su trabajo.*
[Jorge is going to rest when he finishes his work.]
- (2) *Jorge siempre descansa cuando termina su trabajo.*
[Jorge always rests when he finishes his work.]

In contrast, the indicative is used when the event is certain or habitual as in (2). The indicative encodes the certainty of the outcome. The habitual action does not reveal ambiguity or doubt. When Jorge's work is finished, he always rests.

Acquisitional Challenges

The subjunctive/indicative target for the present study was limited to present tense, third person singular *-ar* verbs. The Spanish mood distinction for third-person singular *-ar* verbs is a non-tonic vocalic change: from an *-a* to an *-e*. In addition to perceptual saliency, VanPatten has formulated a number of Input Processing principles that describe and categorize different types of processing problems that learners may encounter when confronted with novel input. Five of those principles apply to the acquisition of this grammatical feature (VanPatten, 2007).

VanPatten's Lexical Preference Principle (P1b) describes the learner's preference for processing lexical items over grammatical features when both describe the same. Using example (1), learners would tend to rely on the future reference (i.e., *va a descansar* [is going to rest]) over processing the subjunctive inflection, the *-e* in *termine* [finishes] to process the uncertainty of the sentence. In other words, learners will tend to interpret the uncertainty of the clause through the periphrastic future rather than the grammatical inflection of the subjunctive in the dependent clause.

Related is the Preference for the Nonredundancy Principle (P1c), in which learners tend to process non-redundant grammatical markers before redundant ones. In the *cuando* plus subjunctive/indicative, the targeted grammatical feature (subjunctive/indicative) is redundant and therefore less attractive and necessary for learners to process. The subjunctive inflection coding is redundant in that it repeats the notion of uncertainty found in the independent clause with the periphrastic future.

It is because learners will tend to process meaningful items before nonmeaningful, as stated in the Meaning before the Nonmeaning Principle (P1d), the grammatical redundant inflection tends to get overlooked. Travis (2003) asserts that in cases like this, the subjunctive "functions as a kind of agreement marker: its meaning is consistent with the proposition expressed, but in itself it does not add any meaning to the construction" (p. 58). The inflectional marking on the subjunctive verb is redundant to the semantic meaning found in the main clause and non-meaningful in the sense that it does

not add to the interpretation of the sentence.

The fourth VanPatten principle that applies to this context is the Availability of Resources Principle (P1e). This principle states that the performance of learners is limited by their processing resources and in order for them to process either redundant or nonmeaningful features, their overall processing resources cannot be drained. The participants in this study are first semester university students and as beginning learners, are often taxed by vocabulary, sentence structure, etc.

Lastly, the Sentence Location Principle (P1f) states that sentence-initial and sentence-final positions are more readily processed than sentence-medial positions. The targeted grammatical feature in this study is not found in an optimal processing position, either first or last, but rather in a medial sentential position.

These factors together make the subjunctive/indicative contrast with *cuando* especially challenging for learners. Despite the challenges inherent in the subjunctive, treatment groups were designed, utilizing PI to push learners to attend to grammatical markers for meaning.

Processing Instruction

Explicit Information

Processing Instruction encourages L2 learners to abandon their current and erroneous processing strategies for strategies that encourage them to make new and accurate form-meaning connections. The two components of PI: Explicit Information (EI) and Structured Input (SI) together accomplish this. Traditional instruction offers grammatical explanations that describe the targeted feature. PI, like traditional instruction does the same, but goes a step further and includes information regarding new strategies that encourage the learner to process the input differently so that their intake is that much richer and correct. The EI of PI encourages the learner to alter their default and often, erroneous processing strategies for new processing strategies that are target language appropriate. PI's EI includes two aspects: 1) explicit information about the target structure, in this case the indicative/subjunctive distinction with *cuando* and 2) explicit information highlighting the new processing strategies for the learner. The new processing strategies will encourage learners to process the target item more effectively, while abandoning their current default strategies.

The EI in this study offered two suggestions to the learners (Appendix A). With regular Spanish verbs, the difference between the present indicative and subjunctive mood forms is a change in a non-tonic vowel (with the exception of 1st person plural, which is tonic), which may be difficult for learners to detect. The EI first directed the learners to detect the subjunctive form itself. A non-tonic vowel in the middle of a sentence is not as salient as in other positions. The second processing problem concerned the subjunctive/indicative distinction. Learners cannot and should not associate the word *cuando*

with the use of the subjunctive or the indicative exclusively. Both moods are possible and their meanings are quite different. Drawing learner's attention to these processing problems also directs learner's attention to the place where the form-meaning connection has to take place. Focusing learner's attention to these processing problems is essential to the PI's EI.

Structured Input

Structured Input (SI) is the practice that PI offers a learner. SI is intentionally and purposefully manipulated to privilege the form within a meaningful context. Structured *input* provides the learner with practice that, interestingly, never requires the learner to produce a form. There are six guidelines for SI practice: 1) use one form at a time, 2) keep meaning in focus, 3) move from sentences to discourses, 4) use both oral and written input, 5) have the learner do something with the input, and 6) keep the learner's processing strategies in mind (Farley, 2005; Lee & VanPatten, 1995; VanPatten, 1996). With the exception of the fifth guideline due to treatment time constraints, all guidelines were followed.

There are two types of SI activities: referential and affective. Both activities push the learner to attend to the target form in order to respond. Referential activities have a correct/incorrect answer if the target form is processed properly. Most often, referential activities take the form of multiple choice questions or item matching. In contrast, affective activities do not have a correct/incorrect response. These activities ask the learner to respond to input that includes the target form with their own opinion or judgment. Theoretically, the learner processes the target form and registers their opinion [see Houston (2010) for an interesting study designed to measure the processing of affective activities]. Affective activities can take formats such as a true/false or yes/no design. Because responses are subjective, all answers are valid. Affective activities serve as a type of input flood in that "they allow learners to see the target forms in meaningful contexts and relate the meaning of each form to their own lives in some way" and obviously provide learners with more "SI with the target feature in the most salient position" (Farley, 2005, p. 87).

Processing Instruction Research

Processing Instruction research began with a study on the acquisition of Spanish object pronouns by native speakers of English. Since that first investigation of VanPatten and Cadierno (1993), much research has been done on PI, expanding the number of L2 languages from Spanish to English (Benati, 2005), French (Lee & Benati, 2007a; VanPatten & Wong, 2004), Italian (Benati, 2001, 2004a, 2004b; Lee & Benati, 2007a, 2007b), and Japanese (Benati, 2008; Lee & Benati, 2007a, 2007b). Not only has PI research encompassed more L2 languages, but also L1 languages. Whereas most PI research has focused on L1 English speakers, other L1 speakers have also been included in PI research.

Native speakers of Chinese and Greek formed Benati's 2005 participants and Lee and Benati (2007a, 2007b) used native speakers of Italian. Benati, Lee, and McNulty (2010) added to the list of native speakers with Cantonese, German, Mandarin, Polish, and Russian.

PI studies began with VanPatten and Cadierno's 1993 work on Spanish object pronouns and was then continued by other researchers (Morgan-Short & Bowden, 2006; Sanz, 2004; VanPatten & Cadierno, 1993; VanPatten & Oikkenon, 1996), and extended to other L2 grammatical features with a variety of inherent acquisitional challenges: Spanish object pronouns, the Spanish preterit tense (Cadierno, 1995), the Spanish preterit/imperfect distinction (Lee, Benati, Aguilar-Sánchez & McNulty, 2007), the Spanish copula (Cheng, 2002, 2004), the Spanish negative informal commands (Lee, Benati, Aguilar-Sánchez, & McNulty, 2007), the Spanish subjunctive (Benati, Lee, & McNulty, 2010; Farley, 2001a, 2001b, 2004a, 2004b; Farley & McCollam, 2004, Henshaw, 2011; Lee & McNulty, 2013; McNulty, 2012), the Spanish anticausative clitic *se* (Toth, 2006), the Italian future tense (Benati, 2001, 2004a), the Italian subjunctive (Lee & Benati, 2007a, 2007b), the Italian gender agreement on adjectives (Benati, 2004b), the French causative (VanPatten & Wong, 2004), the French negative and indefinite article (Wong, 2004b), the French subjunctive (Lee & Benati, 2007a), the Japanese past tense (Lee & Benati, 2007a, 2007b), and the English simple past (Benati, 2005).

PI research has grown in the number of L2 languages, grammatical features with their associated processing problems, and the L1 of the participants. Despite such variety, PI has consistently been shown to be effective in measures of interpretation and production. Some studies have examined the components of PI, namely EI and SI in order to pinpoint the efficacy of PI. Some researchers have concluded that the causative component of PI is SI practice (Benati, 2004a, 2004b; VanPatten & Oikennon, 1996; Wong, 2004b), whereas others found that EI plays a contributing role (Farley, 2004b; Farley & McCollam, 2004).

Of particular interest are two studies: Fernández (2008) and Henry, Culman, and VanPatten (2009) that measured the processing *of* instruction or, in other words, how quickly and accurately instruction was processed, rather than determining the efficacy of the instructional intervention based on the scores of an aural interpretation of written production task. Fernández (2008) used three measures: trials to criterion, response time, and accuracy after criterion for a PI treatment group and SI treatment group. The trials to criterion were defined as the number of SI activities attempted before correctly responding to three consecutive target items and one distractor item in the instructional intervention. Response time was operationalized as a period of time between the stimulus (what was seen or heard) and the selection of the response. The accuracy of criterion figure was calculated by taking the number of correct items after criterion and dividing it by the total number of SI items after the criterion was achieved. For object pronouns, Fernández reported that PI and SI treatment groups behaved equivalently. However, with regard to the subjunctive of doubt,

the PI treatment group that included EI and SI performed better than the SI treatment group. Her results indicate that EI may have a role in processing features and could be dependent on the type of grammatical feature.

Henry, Culman, and VanPatten (2009) partially replicated the Fernández (2008) study but used the grammatical target of German article case markings and only measured the number of trials to criterion. They found that the PI group was significantly quicker in reaching criterion than the SI group, lending support to the notion that EI may be a causative component in processing novel input. These studies prompt this study's research questions: Does the order of introduction of the Explicit Information (before or after Structured Input) on the Spanish subjunctive/indicative contrast after *cuando* affect learner performance as measured by interpretation and form production tasks? Are these effects held over time?

RESEARCH DESIGN

Participants

Volunteer undergraduate student participants were solicited from fourteen intact first year intensive Spanish classes at a large mid-western university. This study is part of a larger study that was comprised of ten treatment groups, only two of which are relevant to the present study. Of the 203 learners who were originally recruited for the ten treatment groups, only fourteen participants completed the study over a five-month period for the two relevant treatment groups. Treatment group ERA (explicit information + referential activities + affective activities) was comprised of seven participants, as was treatment group RAE (referential activities + affective activities + explicit information). The treatment groups differed only when they had access to the EI.

To ensure that participants had the same background knowledge with regard to the Spanish subjunctive/indicative choice with *cuando*, participants had to score below a 50% on the subjunctive items in all four tasks described below. First-year learners were chosen for this study, differentiating the participants in this study from those in other studies (Collentine, 1995, 1997, 1998, 2004; Farley, 2001a, 2001b; Cox & Sanz, 2015; Farley & McCollam, 2004; Gudmestad, 2006; Fernández, 2008; Lubbers Quesada, 1998). Participants also had to participate in all three meetings (pretest, posttest and delayed posttest).

Timeline

The timeline for this study consisted of three meetings with participants as seen in Table 1. The first meeting was conducted in the classroom. Participants were asked to sign an informed consent, complete a background questionnaire, and respond to the tasks described below. The second meeting

included the instructional intervention in a language lab followed by an immediate posttest. The final meeting was in the classroom and participants were asked to take a delayed posttest, approximately five months after the instructional intervention.

Table 1

Timeline of Investigation

<i>When?</i>	<i>What?</i>	<i>Where?</i>
Time 1	Informed Consent Background Questionnaire Pretest	classroom
Time 2	Treatment (ERA, RAE) Posttest	language lab
Time 3	Delayed Posttest	classroom

Note: E=Explicit Information; R=Referential; A=Affective

Instructional Intervention Website

A website was designed using Dreamweaver 9.0, so that multiple instructional treatments could be simultaneously and randomly administered to intact classes. Learners were asked to click the letter on the screen that corresponded with the letter affixed to their workstation. Adjoining workstations were never assigned the same letter. Each letter corresponded to a different instructional intervention. The group ERA received EI (and was available to them throughout the treatment), followed by 24 referential activities and 24 affective activities (Appendices B and C). Treatment group RAE, received the same 24 referential and 24 affective activities as treatment group ERA, but was given access to the EI only after having completed the SI practice. All SI practice items were developed in accordance with VanPatten's SI guidelines and were reviewed by nine near-native and native speakers from a variety of countries to ensure unambiguous SI practice items and answers that were agreed on unanimously.

For both treatment groups, feedback was given as responses were selected. For correct or acceptable answers (in the case of affective activities), students would see: "That's CORRECT! Please select *next question*". For incorrect responses, participants would receive this message: "Remember the form of the verb clues you looked into whether the time is uncertain or not. An *-a* signals certainty (present tense). An *-e* signals uncertainty (future reference). TRY AGAIN".¹

Assessment Tasks

Four assessment tasks were designed for this study measuring sentence interpretation and form production with aural and written prompts as detailed in Table 2 (Appendix D). Most PI investigations tend to use assessment tasks that are similar to Assessments I and IV used in this investigation (e.g., Benati, 2001, 2005; Farley, 2001a; Morgan-Short & Bowden, 2006; VanPatten & Wong, 2004) with aural prompts for interpretation items and written prompts for form production. The present study added Assessments II and III to make the design more balanced. Table 2 shows the breakdown of assessment by aural and written prompt and by what was required of the learner: interpretation or production. This study therefore includes an aural prompt for sentence interpretation *and* form production as well as a written prompt for sentence interpretation *and* form production.

Table 2

Assessment Tasks

	<i>Aural Prompt</i>	<i>Written Prompt</i>
<i>Sentence Interpretation</i>	I	III
<i>Form Production</i>	II	IV

All testing (pretest, posttest, and delayed posttest) were paper and pencil tests as opposed to computerized tests due to logistical considerations and availability of computer labs. Listening tasks were administered before production tasks to eliminate task order effect. A digital recording was made in a soundproof studio by a male, native speaker colleague in an effort to eliminate distracting noises and offer a consistent speech sample. The recording was played only once so that the test would measure real-time comprehension.

In Assessment I, learners heard the first part of a sentence ending in *cuando* and were asked to select the appropriate response. There were six items per assessment task with three items designed to provoke the indicative and three items designed to provoke the subjunctive. All targeted verbs were regular *-ar* verbs. Assessment II also employed an aural prompt. Learners heard the first part of a sentence and then were asked to complete the sentence by correctly conjugating the verb provided. Assessments III and IV were of the same style as Assessments I and II except that instead of an aural prompt, participants were given a written prompt.

RESULTS

Interpretation

Means

As can be seen in Table 3, the interpretation mean scores for total items between Time 1 and Time 2 improved for both treatment groups, although treatment group RAE to a lesser extent. Between Time 2 and Time 3, interpretation mean scores for total items declined for both treatment groups. As Table 3 indicates, the average pretest interpretation score on total items was 6.430 out of a possible 12.000 (54%) for the treatment group ERA and 6.860 out of 12.000 (57%) for the RAE treatment. The average posttest interpretation score on total items for treatment group ERA was 9.430 (79%) and 7.000 (58%) for the RAE treatment. Five months later, the average delayed posttest interpretation score on total items for treatment group ERA was 6.860 (57%) and 5.290 (44%) for the RAE treatment.

Table 3

*Mean Scores for Interpretation Tasks (Descriptive)
for Treatments ERA and RAE*

		<i>Group ERA</i>			<i>Group RAE</i>		
		<i>M</i>	<i>N</i>	<i>SD</i>	<i>M</i>	<i>N</i>	<i>SD</i>
<i>Pretest (Time 1)</i>	Total	6.43	7	1.618	6.86	7	2.035
	Subjunctive	2.29	7	0.951	2.86	7	1.345
	Indicative	4.14	7	1.069	4.00	7	1.000
<i>Posttest (Time 2)</i>	Total	9.43	7	2.149	7.00	7	4.203
	Subjunctive	4.57	7	1.134	3.29	7	2.430
	Indicative	4.86	7	1.215	3.71	7	2.059
<i>Delayed Posttest (Time 3)</i>	Total	6.86	7	0.690	5.29	7	1.976
	Subjunctive	3.43	7	0.976	2.86	7	0.900
	Indicative	3.43	7	0.976	2.43	7	1.397

Note: E=Explicit Information; R=Referential; A=Affective

The interpretation mean scores for subjunctive items between Time 1 and Time 2 improved for both treatment groups as can be seen in Table 3. Between Time 2 and Time 3, interpretation mean scores for subjunctive items also declined for both treatment groups. The average pretest interpretation score on subjunctive items was 2.290 out of 6.000 (38%) for the ERA treatment group and 2.860 out of 6.000 (48%) for the RAE treatment group. The average posttest interpretation score on subjunctive items for treatment group ERA was 4.570

(76%) and 3.290 (55%) for the RAE treatment. Five months later, the average delayed posttest interpretation score on subjunctive items for treatment group ERA was 3.430 (57%) and 2.860 (48%) for the RAE treatment.

As can be seen in Table 3, the interpretation mean scores for indicative items between Time 1 and Time 2 improved for treatment group ERA and declined for treatment group RAE. Between Time 2 and Time 3, interpretation mean scores for indicative items declined for both treatment groups. The average pretest interpretation score on indicative items was 4.140 out of 6.000 (69%) for the ERA treatment group and 4.000 out of 6.000 (67%) for the RAE treatment group. The average posttest interpretation score on indicative items for ERA treatment group was 4.860 (81%) and 3.710 (62%) for the RAE treatment group. Five months later, the average delayed posttest interpretation score on indicative items for treatment group ERA was 3.430 (57%) and 2.430 (41%) for the RAE treatment.

Pretest Scores

A one-way ANOVA performed on interpretation pretest scores revealed no significant differences between the two treatment groups for total items, $F(1, 12) = 0.190, p = .671$. Additionally, one-way ANOVAs were run on the interpretation scores of subjunctive items ($F(1, 12) = 0.842, p = .377$) and the interpretation scores on indicative items ($F(1, 12) = 0.067, p = .801$) individually to further establish that both treatment groups started the experiment with equivalent knowledge of the target structure. Therefore, any differences found on subsequent scores between treatment groups can be attributed to treatment effects rather than to preexisting differences. Table 4 presents a summary of the F statistic and *p* values that resulted from the one-way ANOVAs.

Table 4

One-way ANOVA on Pretest scores for Treatments ERA and RAE

		<i>df</i>	<i>F</i>	<i>p value</i>
<i>Interpretation</i>	Total	1, 12	0.190	.671
	Subjunctive	1, 12	0.842	.377
	Indicative	1, 12	0.067	.801

Note: E=Explicit Information; R=Referential; A=Affective.

Significant at the $p < .05$ level.

Repeated Measures ANOVA

Having determined that there were no statistical differences among pretest scores, the interpretation score for total items was submitted first to a repeatedmeasures ANOVA. In order to clarify or pinpoint the component

responsible for the significance or lack thereof, separate repeated measures ANOVAs were run for the interpretation score on subjunctive items and the interpretation score on indicative items. Table 5 presents this data.

Table 5

ANOVA on Location of Explanation Comparison of Treatments ERA and RAE for Interpretation Measures

		Total	Sub.	Ind.
<i>Time</i>	<i>df</i>	2, 11	2, 11	2, 24
	<i>F</i>	2.054	6.604	4.742
	<i>p</i> value	0.175	0.013*	0.018*
<i>Treatment</i>	<i>df</i>	1, 12	1, 12	1, 12
	<i>F</i>	2.500	1.130	2.723
	<i>p</i> value	0.140	0.309	0.125
<i>Time * Treatment</i>	<i>df</i>	2, 11	2, 11	2, 24
	<i>F</i>	3.513	3.712	0.652
	<i>p</i> value	0.066	0.059	0.530

Note: E=Explicit Information; R=Referential; A=Affective; Sub=Subjunctive; Ind=Indicative.

* $p < .05$.

The pretest, posttest, and delayed posttest scores were submitted to a 2 x 3 repeated measures ANOVA for the interpretation score on total items. Treatment (ERA or RAE) was the between-subjects factor and the repeated measure, or the within subjects factor, was Time (pretest, posttest, delayed posttest). The results indicate that there was not a significant main effect for Time ($F(2, 11) = 2.054, p = .175$), no significant main effect for Treatment ($F(1, 12) = 2.500, p = .140$) and no significant interaction ($F(2, 11) = 3.513, p = .066$) at the alpha = .05 level. These results indicate that neither treatment groups improved significantly on the interpreting sentences as measured by a total score as a result of instruction, and that both treatment groups behaved similarly.

The pretest, posttest, and delayed posttest scores were submitted to a 2 x 3 repeated measures ANOVA for the interpretation score on subjunctive items. Treatment (ERA or RAE) was the between-subjects factor and the repeated measure, or the within subjects factor, was Time (pretest, posttest, delayed posttest). The results indicate that there was a significant main effect for Time ($F(2, 11) = 6.604, p = .013$), no significant main effect for Treatment ($F(1, 12) = 1.130, p = .309$) and no significant interaction ($F(2, 11) = 3.712, p = .059$) at the alpha = .05 level. These results indicate that both treatment groups improved significantly on the interpreting sentences as measured by a subjunctive score as a result of instruction, and that both treatment groups behaved similarly.

The pretest, posttest, and delayed posttest scores were submitted to a 2 x 3 repeated measures ANOVA for the interpretation score on indicative items. Treatment (ERA or RAE) was the between-subjects factor and the repeated

measure, or the within subjects factor, was Time (pretest, posttest, delayed posttest). The results indicate that there was a significant main effect for Time ($F(2, 24) = 4.742, p = .018$), no significant main effect for Treatment ($F(1, 12) = 2.723, p = .125$) and no significant interaction ($F(2, 24) = 0.652, p = .530$) at the alpha .05 level. These results indicate that both treatment groups improved significantly on the interpreting sentences as measured by an indicative score as a result of instruction, and that both treatment groups behaved similarly.

Post hoc Analysis

Interpretation scores on total items. Although there was not a significant interaction between Time and Treatment ($F(2, 11) = 3.513, p = .066$) for interpretation scores on total items, because the p value was approaching significance, data for each individual treatment group (ERA, RAE) was submitted to an individual repeated measures ANOVA to investigate the trends across time for each treatment, as seen in Table 6.

Table 6
ANOVA for P values for Time * Treatment that Approach Significance on Interpretation Scores of Total Items

	Time		
	df	F	p value
ERA	2, 12	6.809	.011*
RAE	2, 12	0.786	.478

Note: E=Explicit Information; R=Referential; A=Affective.
* $p < .05$.

The results indicate that for treatment group ERA, there was a significant main effect for Time ($F(2, 12) = 6.809, p = .011$). In order to tease apart this finding, a pairwise comparison was done to pinpoint where the differences were, as noted in Table 7.

Table 7
 P values from Pairwise Comparisons for Treatment ERA on Interpretation Scores on Total Items

	Time 1	Time 2	Time 3
Time 1	X	.014*	.555
Time 2	-	X	.049*
Time 3	-	-	X

Note: E=Explicit Information; R=Referential; A=Affective.
* $p < .05$.

As demonstrated in Table 7, interpretation scores on total items for treatment group ERA has a significant difference between Time 1 and Time 2 (pretest to posttest) as well as Time 2 and Time 3 (posttest to delayed posttest), but is not significant between Time 1 and Time 3 (pretest to delayed posttest).

Interpretation scores on subjunctive items. Although there was not a significant interaction between Time and Treatment ($F(2, 11) = 3.712, p = .059$) for interpretation scores on subjunctive items, because the p value was approaching significance, data for each individual treatment group (ERA, RAE) was submitted to an individual repeated measures ANOVA to investigate the trends across time for each treatment, as seen in Table 8.

Table 8
*ANOVA for P values for Time * Treatment that Approach Significance on Interpretation Scores of Subjunctive Items*

	Time		
	df	F	p value
ERA	2, 12	7.291	.008*
RAE	2, 12	0.149	.863

Note: E=Explicit Information; R=Referential; A=Affective.

* $p < .05$.

The results indicate that for treatment group ERA, there was a significant main effect for Time ($F(2, 12) = 7.291, p = .008$). In order to tease apart this finding, a pairwise comparison was done to pinpoint where the differences were, as shown in Table 9.

Table 9
P values from Pairwise Comparisons for Treatment ERA on Interpretation Scores for Subjunctive Items

	Time 1	Time 2	Time 3
Time 1	X	.005*	.066
Time 2	-	X	.172
Time 3	-	-	X

Note: E=Explicit Information; R=Referential; A=Affective.

* $p < .05$.

As demonstrated in Table 9, interpretation scores on subjunctive items for treatment group ERA demonstrate a significant difference between Times 1 and 2 (pretest to posttest) but not between Times 2 and 3 (posttest to delayed posttest), and are also not significant between Times 1 and 3 (pretest to delayed posttest).

Production

Means

As can be seen in Table 10, the production mean scores for total items between Time 1 and Time 2 improved for treatment group ERA, but declined for treatment group RAE. Between Time 2 and Time 3, production mean scores for total items declined for treatment group ERA and improved for treatment group RAE. As Table 10 indicates, the average pretest production score on total items was 5.710 out of a possible 12.000 (48%) for the treatment group ERA and 6.140 out of 12.000 (51%) for the RAE treatment. The average posttest production score on total items for treatment group ERA was 8.290 (69%) and 5.290 (44%) for the RAE treatment. Five months later, the average delayed posttest production score on total items for treatment group ERA was 4.860 (41%) and 6.000 (50%) for the RAE treatment.

The production mean scores for subjunctive items between Time 1 and Time 2 improved for both treatment groups as can be seen in Table 10. Between Time 2 and Time 3, production mean scores for subjunctive items also declined for both treatment groups. The average pretest production score on subjunctive items was 0.000 out of 6.000 (0%) for the treatment group ERA and 1.000 out of 6.000 (17%) for the RAE treatment group. The average posttest production score on subjunctive items for treatment group ERA was 3.000 (50%) and 1.860 (31%) for the RAE treatment. Five months later, the average delayed posttest production score on subjunctive items for treatment group ERA was 0.000 (0%) and 0.430 (7%) for the RAE treatment.

As shown in Table 10, the production mean scores for indicative items between Time 1 and Time 2 declined for both treatment groups. Between Time 2 and Time 3, production mean scores for indicative items declined for treatment group ERA and improved for treatment group RAE. The average pretest production score on indicative items was 5.710 out of 6.000 (95%) for the treatment group ERA and 5.140 out of 6.000 (86%) for the RAE treatment group. The average posttest production score on indicative items for treatment group ERA was 5.290 (88%) and 3.430 (57%) for the RAE treatment. Five months later, the average delayed posttest production score on indicative items for treatment group ERA was 4.860 (81%) and 5.570 (93%) for the RAE treatment.

Table 10
Mean Scores for Production Tasks (Descriptive)
for Treatments ERA and RAE

		<i>Group ERA</i>			<i>Group RAE</i>		
		M	N	SD	M	N	SD
<i>Pretest</i> <i>(Time 1)</i>	Total	5.71	7	0.488	6.14	7	1.464
	Subjunctive	0.00	7	0.000	1.00	7	1.414
	Indicative	5.71	7	0.488	5.14	7	1.864
<i>Posttest</i> <i>(Time 2)</i>	Total	8.29	7	2.059	5.29	7	3.352
	Subjunctive	3.00	7	1.915	1.86	7	2.116
	Indicative	5.29	7	0.951	3.43	7	1.988
<i>Delayed</i> <i>Posttest</i> <i>(Time 3)</i>	Total	4.86	7	1.215	6.00	7	1.528
	Subjunctive	0.00	7	0.000	0.43	7	1.134
	Indicative	4.86	7	1.215	5.57	7	0.787

Note: E=Explicit Information; R=Referential; A=Affective

Pretest Scores

A one-way ANOVA performed on production pretest scores revealed no significant differences between the two treatment groups for total items, $F(1, 12) = 0.540$, $p = .477$. Additionally, one-way ANOVAs were run on the production scores of subjunctive items ($F(1, 12) = 3.500$, $p = .477$) and the production scores on indicative items ($F(1, 12) = 0.615$, $p = .448$) individually to further establish that both treatment groups started the experiment with equivalent knowledge of the target structure. Therefore, any differences found on subsequent scores between treatment groups can be attributed to treatment effects rather than to preexisting differences. Table 11 presents a summary of the F statistic and p values that resulted from the one-way ANOVAs.

Table 11
One-way ANOVA on Pretest scores for Treatments ERA and RAE

		<i>df</i>	<i>F</i>	<i>p value</i>
<i>Production</i>	Total	1, 12	0.540	.477
	Subjunctive	1, 12	3.500	.086
	Indicative	1, 12	0.615	.448

Note. E=Explicit Information; R=Referential; A=Affective.

* $p < .05$.

Repeated Measures ANOVA

Having determined that there were no statistical differences among pretest scores, the production score for total items was submitted first to an individual repeated measures ANOVA. In order to clarify or pinpoint the component responsible for the significance or lack thereof, separate repeated measures ANOVAs were run for the production score on subjunctive items and the production score on indicative items. Table 12 presents this data.

Table 12

ANOVA on Location of Explanation Comparison of Treatments ERA and RAE for Production Measures

		Total	Sub.	Ind.
Time	df	2, 11	2, 11	2, 24
	F	1.880	12.610	2.377
	p value	0.199	0.001*	0.114
Treatment	df	1, 12	1, 12	1, 12
	F	0.392	0.027	3.429
	p value	0.543	0.872	0.162
Time * Treatment	df	2, 11	2, 11	2, 24
	F	3.757	2.355	3.057
	p value	0.057	0.141	0.066

Note: E=Explicit Information; R=Referential; A=Affective; Sub=Subjunctive; Ind=Indicative.

* $p < .05$.

The pretest, posttest, and delayed posttest scores were submitted to a 2 x 3 repeated measures ANOVA for the production score on total items. Treatment (ERA or RAE) was the between-subjects factor and the repeated measure, or the within subjects factor, was Time (pretest, posttest, delayed posttest). The results indicate that there was not a significant main effect for Time ($F(2, 11) = 1.880, p = .199$), no significant main effect for Treatment ($F(1, 12) = 0.392, p = .543$), and no significant interaction ($F(2, 11) = 3.757, p = .057$) at the alpha = .05 level. These results indicate that neither treatment groups improved significantly on form production as measured by a total score as a result of instruction, and that both treatment groups behaved similarly.

The pretest, posttest, and delayed posttest scores were submitted to a 2 x 3 repeated measures ANOVA for the production score on subjunctive items. Treatment (ERA or RAE) was the between-subjects factor and the repeated measure, or the within subjects factor, was Time (pretest, posttest, delayed posttest). The results indicate that there was a significant main effect for Time ($F(2, 11) = 12.610, p = .001$), no significant main effect for Treatment ($F(1, 12)$

= 0.027, $p = .872$), and no significant interaction ($F(2, 11) = 2.355, p = .141$) at the $\alpha = .05$ level. These results indicate that both treatment groups improved significantly on subjunctive form production as measured by the subjunctive score as a result of instruction, and that both treatment groups behaved similarly.

The pretest, posttest, and delayed posttest scores were submitted to a 2 x 3 repeated measures ANOVA for the production score on indicative items. Treatment (ERA or RAE) was the between-subjects factor and the repeated measure, or the within subjects factor, was Time (pretest, posttest, delayed posttest). The results indicate that there was not a significant main effect for Time ($F(2, 24) = 2.377, p = .114$), no significant main effect for Treatment ($F(1, 12) = 3.429, p = .162$), and no significant interaction ($F(2, 24) = 3.057, p = .066$) at the $\alpha = .05$ level. These results indicate that neither treatment groups improved significantly on indicative form production as measured by the indicative score as a result of instruction, and that both treatment groups behaved similarly.

Post hoc Analysis

Production scores on total items. Although there was not a significant interaction between Time and Treatment ($F(2, 11) = 3.757, p = .057$) for production scores on total items, because the p value was approaching significance, data for each individual treatment group (ERA, RAE) was submitted to an individual repeated measures ANOVA to investigate the trends across time for each treatment, as seen in Table 13.

Table 13

*ANOVA for p values for Time * Treatment that approach significance on Production Scores of Total Items*

	Time		
	<i>df</i>	<i>F</i>	<i>p</i> value
<i>ERA</i>	2, 12	4.528	.075
<i>RAE</i>	2, 12	0.583	.573

Note: E=Explicit Information; R=Referential; A=Affective.

Significant at the $p < .05$ level.

The results indicate that there was not a significant main effect for Time for treatment group ERA or RAE, therefore a pairwise comparison was not required.

Production scores on indicative items. Although there was not a significant interaction between Time and Treatment ($F(2, 24) = 3.057, p = .066$) for production scores on indicative items, because the p value was approaching significance, data for each individual treatment group (ERA, RAE) was

submitted to an individual repeated measures ANOVA to investigate the trends across time for each treatment, as seen in Table 14.

Table 14

*ANOVA for P values for Time * Treatment that Approach Significance
Production Scores of Indicative Items*

	Time		
	<i>df</i>	<i>F</i>	<i>p</i> value
<i>ERA</i>	2, 12	1.528	.256
<i>RAE</i>	2, 12	3.057	.085

Note: E=Explicit Information; R=Referential; A=Affective.
Significant at the $p < .05$ level.

The results indicate that there was not a significant main effect for Time for treatment group ERA or RAE and therefore a pairwise comparison was not required.

DISCUSSION/CONCLUSION

Knowing that some researchers (Cox & Sanz, 2015; Farley, 2004b; Farley & McCollam, 2004; Fernández, 2008; Henry, Culman & VanPatten, 2009) have found that Explicit Information contributes positively to learner performance, the question became whether the timing of when learners were introduced to the EI had any effect. Does the location of the Explicit Information (before or after Structured Input) on the Spanish subjunctive/indicative contrast after *cuando* affect learner performance as measured by interpretation and form production tasks?

There was a significant main effect for Time on three scores: the interpretation scores on subjunctive and indicative items and the production scores on subjunctive items. The main effect for Time for these three scores is attributable to the differences between pretest and posttest scores. In other words, the instructional intervention affected these three scores as one might expect, but did so similarly between treatments. As there was no main effect for Treatment and no significant interaction between Time and Treatment, the simple answer to the research question is that the location of the Explicit Information (before or after Structured Input) does not affect learner performance on interpretation and production tasks. The timing of the EI, before or after SI, did not change learner outcome between treatments. It should be noted that the SI practice for both treatments did offer metalinguistic information about the grammatical feature. Although this explicit feedback cannot be considered EI, it did provide all learners with processing information. Removing this explicit feedback or simplifying this feedback to correct/incorrect in the future might provide a clearer result.

Revisiting the repeated measures ANOVAs (see Tables 5 and 12), there were four scores that approached significance and were considered further (see Tables 6-9, 13, 14). For the four scores that approached significance, interpretation scores on total and subjunctive items as well as production scores on total and indicative items, only treatment group ERA had a main effect for Time in the two interpretation scores.

The ERA treatment group demonstrated a benefit from pretest to posttest on interpretation scores on total items and a significant change from posttest to delayed posttest, but the gains in the posttest were not sustained five months later on the delayed posttest. ERA's interpretation scores on subjunctive items improved from pretest to posttest, but gains were not sustained through the delayed posttest. There was no main effect for Time for the treatment group RAE.

Group ERA received the Explicit Information in advance of the Structured Input practice and the metalinguistic feedback from the referential activities confirmed their knowledge. Theoretically, group RAE received metalinguistic knowledge via the SI practice and were able to confirm their hypothesis about the grammatical feature with the EI. Given that statistically each treatment behaved equivalently, it would seem that the placement of EI is not critical. Whereas it cannot be concluded that the location or timing of the EI is statistically significant, there are indications that the metalinguistic feedback offered by the referential activities may have contributed to or compensated for the absence of the EI first in group RAE. Although these results are not statistically significant, this study contributes to the field of knowledge by ruling out order of EI and pointing to explicit feedback as a possible causative or contributive component to the success of Processing Instruction. However, given the small number of participants in this study and the mixed results that prior research has demonstrated on the importance of EI, replication of this study or studies with a different grammatical focus could help to better define EI's role in PI acquisition.

NOTE

1. It should be noted that this explicit feedback was provided to both treatment groups (ERA and RAE) and, although not considered EI, did contain processing information.

REFERENCES

- Benati, A. (2001). A comparative study of the effects of processing instruction and output-based instruction on the acquisition of the Italian future tense. *Language Teaching Research*, 5, 95-127.
- Benati, A. (2004a). The effects of structured input and explicit information on the acquisition of Italian future tense. In B. VanPatten (Ed.), *Processing instruction: Theory, research, and commentary* (pp. 207-255). Mahwah, NJ: Erlbaum.
- Benati, A. (2004b). The effects of processing instruction and its components on the acquisition of gender agreement in Italian. *Language Awareness*, 13, 67-80.
- Benati, A. (2005). The effects of processing instruction, traditional instruction and meaning-output based instruction on the acquisition of English past simple tense. *Language Teaching Research*, 9(1), 67-93.
- Benati, A. (2008). *Japanese language teaching: A communicative approach*. London: Continuum.
- Benati, A., Lee, J. F., & McNulty, E. M. (2010). Exploring the effects of processing instruction on a discourse-level guided composition with the Spanish subjunctive after the adverb *cuando*. In A. Benati, & J. F. Lee (Eds.) *Processing instruction and discourse* (pp. 97-147). London: Continuum.
- Cadierno, T. (1995). Formal instruction from a processing perspective: An investigation into the Spanish past tense. *The Modern Language Journal*, 79, 179-93.
- Cheng, A. (2002). The effects of processing instruction on the acquisition of *ser* and *estar*. *Hispania*, 85, 308-323.
- Cheng, A. (2004). Processing instruction and Spanish *ser* and *estar*: Forms with semantic-aspectual value. In B. VanPatten (Ed.), *Processing instruction: Theory, research, and commentary* (pp. 119-141). Mahwah, NJ: Erlbaum.
- Collentine, J. (1995). The development of complex syntax and mood-selection abilities by intermediate-level learners of Spanish. *Hispania*, 78, 122-135.
- Collentine, J. (1997). The effects of irregular stems on the detection of verbs in the subjunctive. *Spanish Applied Linguistics*, 1, 3-23.
- Collentine, J. (1998). Processing instruction and the subjunctive. *Hispania*, 81, 576-587.
- Collentine, J. (2004). Commentary: Where PI research has been and where it should be going. In B. VanPatten (Ed.), *Processing instruction: Theory, research, and commentary* (pp. 119-141). Mahwah, NJ: Erlbaum.
- Cox, J. & Sanz, C. (2015). Deconstructing PI for the ages: Explicit instruction vs. practice young and older adult bilinguals. *IRAL*, 53(2), 225-248.

- Farley, A. P. (2001a). Processing instruction and meaning-based output instruction: A comparative study. *Spanish Applied Linguistics*, 5, 57-94.
- Farley, A. P. (2001b). Authentic processing instruction and the Spanish subjunctive. *Hispania*, 84, 289-299.
- Farley, A. P. (2004a). The relative effects of processing instruction and meaning-based output instruction. In B. VanPatten (Ed.), *Processing instruction: Theory, research, and commentary* (pp. 143-168). Mahwah, NJ: Erlbaum.
- Farley, A. P. (2004b). Processing instruction and the Spanish subjunctive: Is explicit information needed? In B. VanPatten (Ed.), *Processing instruction: Theory, research, and commentary* (pp. 227-239). Mahwah, NJ: Erlbaum.
- Farley, A. P. (2005). *Structured input: Grammar instruction for the acquisition-oriented classroom*. New York, NY: McGraw-Hill.
- Farley, A. P., & McCollam, K. (2004). Learner readiness and L2 production in Spanish: Processability theory on trial. *Estudios de Lingüística Aplicada*, 22(40), 47-69.
- Fernández, C. (2008). Reexamining the role of explicit information in processing instruction. *Studies in Second Language Acquisition*, 30, 277-305.
- Gudmestad, A. (2006). L2 variation and the Spanish subjunctive: Linguistic features predicting use. In C. Klee, & T. Face (Eds.), *Selected papers of the 7th conference on the acquisition of Spanish and Portuguese as first and second languages* (pp. 170-184). Somerville, MA: Cascadilla Press.
- Henry, N., Culman, H., & VanPatten, B. (2009). More on the effects of explicit information in structured SLA: A partial replication and a response to Fernández (2008). *Studies in Second Language Acquisition*, 31, 559-575.
- Henshaw, F. (2011). How effective are affective activities? Relative benefits of two types of structured input activities as part of a computer-delivered lesson on the Spanish subjunctive. *Language Teaching Research*, 16(3) 393-414.
- Houston, T. (2010). Affective structured input online: How authentic are learner responses? *Hispania*, 93(2), 218-234.
- Lee, J. F., & Benati, A. (2007a). *Delivering processing instruction in classrooms and virtual contexts: Research and practice*. London: Equinox.
- Lee, J. F., & Benati, A. (2007b). *Second language processing: An analysis of theory, problems and solutions*. London: Continuum.

- Lee, J. F., Benati, A., Aguilar-Sánchez, J., & McNulty, E. M. (2007). Comparing three modes of delivering processing instruction on preterite/imperfect distinction and negative informal commands in Spanish. In J. F. Lee, & A. Benati (Eds.) *Delivering processing instruction in classrooms and virtual contexts: Research and practice* (pp. 73-98). London: Equinox.
- Lee, J. F., & McNulty, E. M. (2013). The effects of language background on the results of processing instruction on the Spanish subjunctive/indicative contrast after the adverb *cuando*. In J. F. Lee, & A. Benati (Eds.), *Individual differences and processing instruction*, (pp. 49-81). London: Equinox.
- Lee, J. F., & VanPatten, B. (1995). *Making communicative language teaching happen*. New York, NY: McGraw-Hill.
- Lubbers Quesada, M. (1998). L2 acquisition of the Spanish subjunctive mood and prototype schema development. *Spanish Applied Linguistics*, 2(1), 1-23.
- McNulty, E. M. (2012). *On activity types and activity order in the processing instruction of Spanish cuando plus subjunctive/indicative* (doctoral dissertation). Indiana University, Bloomington, Indiana.
- Morgan-Short, K., & Bowden, H. W. (2006). Processing instruction and meaningful output-based instruction: Effects on second language development. *Studies in Second Language Acquisition*, 28, 31-65.
- Sanz, C. (2004). Computer delivered implicit versus explicit feedback in processing instruction. In B. VanPatten (Ed.), *Processing instruction: Theory, research, and commentary* (pp. 241-255). Mahwah, NJ: Erlbaum.
- Toth, P. (2006). Processing instruction and a role for output in second language acquisition. *Language Learning*, 56(2), 319-385.
- Travis, C. (2003). The semantics of the Spanish subjunctive: Its use in the natural semantic metalanguage. *Cognitive Linguistics*, 14(1), 47-69.
- VanPatten, B. (1996). *Input processing and grammar instruction: Theory and research*. Norwood, NJ: Ablex.
- VanPatten, B. (2007). Input processing in adult second language acquisition. In B. VanPatten, and J. Williams (Eds.), *Theories in second language acquisition*, (pp. 115-135). Mahwah, NJ: Erlbaum.
- VanPatten, B., & Cadierno, T. (1993). Explicit instruction and input processing. *Studies in Second Language Acquisition*, 15, 225-243.
- VanPatten, B., & Oikennon, S. (1996). Explanation vs. structured input in processing instruction. *Studies in Second Language Acquisition*, 18, 495-510.
- VanPatten, B., & Wong, W. (2004). Processing instruction and the French causative: Another replication. In B. VanPatten (Ed.), *Processing instruction: Theory, research, and commentary* (pp. 97-118). Mahwah, NJ: Erlbaum.

- Wong, W. (2004). Processing instruction in French: The roles of explicit information and structured input. In B. VanPatten (Ed.), *Processing instruction: Theory, research, and commentary* (pp. 187-205). Mahwah, NJ: Erlbaum.

APPENDIX A

Processing Instruction's Explicit Information

Quando: Subjunctive or Indicative?

What is tricky about the time conjunction, *cuando* (when), is that it can be used with both indicative and subjunctive verb forms. Some other conjunctions of time work in the same way.

en cuanto	as soon as
hasta que	until
tan pronto como	as soon as

Today, however, we will focus only on *cuando*.

Subjunctive...when we are not certain of the outcome

We use the present subjunctive with *cuando* when we want to refer to something that hasn't happened yet or is anticipated. It can also be used when it refers to something that you aren't sure when or even if it will happen.

Actually, when a sentence contains a future reference, by its very nature, uncertain -- the verb following *cuando* will be in the subjunctive! Keep in mind that future can be expressed in a number of ways in Spanish, just like English. You have already learned two different ways of expressing the future in Spanish.

Spanish uses the periphrastic future (*ir + a + infinitive*) to express the future:

CUANDO + Subjunctive	FUTURE reference
Quando le toque la lotería,	va a comprar una casa.
When he wins the lottery,	he is going to buy a house.

Does he know **when** he will win the lottery? Better yet, does he even know **if** he will win the lottery? Of course not! Spanish uses the subjunctive mood to emphasize the uncertainty of the event.

The simple present tense can also express the future, especially with adverbs (tomorrow, next week, this Friday, etc.):

FUTURE reference	CUANDO + Subjunctive
Mañana Jorge descansa	cuando termine su trabajo.
Tomorrow Jorge (will) rest	when he finishes his work.

Will Jorge rest at 9:00 am? 2:00 pm? We don't know. We only know that it will be tomorrow, but we are uncertain of the time. The event of finishing his work hasn't happened yet. The subjunctive indicates that

*****OJO***** *Did you notice that cuando can come at the beginning or in the middle of a sentence?*

the event hasn't taken place yet.

Indicative...when we are certain of the outcome

In contrast, we use the indicative with *cuando* when the sentence refers to something that habitually happens or something that happened in the past.

For the next example, we will focus on when the present tense encodes something that happens habitually. Just like English, adverbs (like generally, always etc.) are sometimes used with the present tense to indicate a habitual action. Notice how the following example in the present tense shows a habitual action.

HABITUAL reference

Siempre David me llama
David always calls me

CUANDO + Indicative (present)

cuando necesita algo.
when he needs something.

When does David call me? ...when he needs something. He has the habit of calling me each and every time he needs something.

What are the subjunctive forms?

So, how do we recognize the subjunctive when we hear it or see it? To form the subjunctive:

1. start with the *yo* form of the verb in the present tense;
2. take off the *-o* or *-oy*; and
3. change the endings. Ø
 1. *-ar* verbs take *-er/-ir* verb endings, and
 2. *-er/-ir* verbs take *-ar* verb endings

-ar verbs

As the following examples show with *-ar* verbs, the third person singular indicative (present) form ends in an *-a*, whereas the third person subjunctive form ends in an *-e*.

Take, for example *hablar*. Following the steps listed above,

1. hablo (*yo* form of *hablar*)
2. habl- (taking off the *-o*)
3. hable (adding the *-er/-ir* verb ending)

Infinitive	Indicative	Subjunctive
visitar	visita	visite
nadar	nada	nade

¡OJO!



What is challenging about these subjunctive forms is that the difference between an *-a* and an *-e* may be difficult to detect (or perceive). Detecting the difference is even more difficult because the stressed syllable is not the one where the *-a* or *-e* is located! Native speakers of English have to train their ears to detect this difference. We want you to learn to detect that the *-e* at the end of an *-ar* verb when it follows *cuando* refers to an unspecified time: the subjunctive.

In the practice activities that follow, we are going to focus only on *-ar* verbs, meaning that you will be learning to detect a sound at the end of the verb: the *-e* of the subjunctive and the *-a* of the indicative.

The *-e* of the subjunctive will tell you that the time an event takes place is uncertain or that it hasn't happened yet.

The *-a* of the indicative will tell you that an event occurs regularly or habitually.

APPENDIX B

Examples of Referential Activities

Choose the correct interpretation for each sentence. Remember, the form of the verb clues you into whether the time is uncertain or not. Since all the verbs are *-ar* verbs, an *-e* signals uncertainty whereas an *-a* signals certainty.

1. Cuando Juan toca la guitarra...
 - a. We don't know when Juan will play the guitar.
 - b. Juan plays the guitar all the time.

2. Cuando Rosa cante con música...
 - a. We don't know when Rosa will sing with music.
 - b. Rosa always sings with music.

Similar to the previous activity, you will now hear part of a sentence. Choose the correct interpretation for each sentence. Remember, the form of the verb clues you in to whether the time is uncertain or not. Since all the verbs are *-ar* verbs, an *-e* signals uncertainty whereas an *-a* signals certainty.

1. Cuando el profesor explica la tarea...
 - a. We don't know when the professor will explain the homework.
 - b. The professor normally explains the homework.

2. Cuando el estudiante ande por el campus...
 - a. We don't know when the student will walk on campus.
 - b. The student walks on campus all the time.

For each item indicate whether it is uncertain when the action mentioned will take place or if it takes place habitually. As all the verbs are *-ar* verbs, uncertainty is signaled with an *-e* and habitual action with an *-a*.

- | | uncertain | habitual |
|---------------------------------------|--------------------------|--------------------------|
| 1. Cuando Eli baila salsa,... | <input type="checkbox"/> | <input type="checkbox"/> |
| 2. Cuando Paco llame a sus padres,... | <input type="checkbox"/> | <input type="checkbox"/> |

Similar to the previous activity, you will now hear part of a sentence. For each item indicate whether it is uncertain when the action mentioned is going to take place or if it takes place habitually. As all the verbs are *-ar* verbs, uncertainty is signaled with an *-e* and habitual action with an *-a*.

- | | uncertain | habitual |
|---|--------------------------|--------------------------|
| 1. Cuando Virginia baja del autobús,... | <input type="checkbox"/> | <input type="checkbox"/> |
| 2. Cuando Lidia compre mucho,... | <input type="checkbox"/> | <input type="checkbox"/> |

Complete the sentence by choosing the appropriate ending. Keep in mind, that all the verbs used with *cuando* are *-ar* verbs.

Remember, uncertain conditions have a verb with a future orientation and are signaled with a verb ending in an *-e*. Habitual conditions, on the other hand, have verbs in the present tense, usually with an adverb and the verbs end in an *-a*.

1. Cuando Marta cocine bien,
 - a. va a comer sano.
 - b. come sano.

2. Cuando Diego practica el deporte, normalmente
 - a. va a conocer más gente.
 - b. conoce más gente.

Just like the activity you just completed, you will hear the first part of a sentence. Complete the sentence by choosing the appropriate ending. Keep in mind that all the verbs used with *cuando* are *-ar* verbs.

Remember, uncertain conditions have a verb with a future orientation and are signaled with a verb ending in an *-e*. Habitual conditions, on the other hand, have verbs in the present tense, usually with an adverb, and the verbs end in an *-a*.

1. Cuando Marta estudie mucho,
 - a. va a salir bien en el examen.
 - b. sale bien en el examen.

2. Cuando Diego espera para el autobús, generalmente
 - a. va a leer un libro.
 - b. lee un libro.

APPENDIX C**Examples of Affective Activities**

Indicate whether, in your opinion, your instructor will do the following things. Keep in mind that the form of the verbs following *cuando* in 1-6 encodes the uncertainty of the time. sí no

1. Cuando esté de vacaciones, tu instructor(a) va a tener más tiempo.
2. Cuando compre zapatos deportivos nuevos, tu instructor(a) va a hacer más ejercicio.

Similar to the previous activity, you will now hear sentences. Indicate whether, in your opinion, your instructor will do the following things. Keep in mind that the form of the verbs following *cuando* in 1-6 encodes the uncertainty of the time. sí no

1. Cuando llame por teléfono, tu instructor(a) va a charlar con su hermana.
2. Cuando cocine algo elegante, tu instructor(a) va a tener invitados.

Indicate whether, in your opinion, your instructor regularly does the following things. Keep in mind that the form of the verbs following *cuando* in 13-24 encodes the habitualness of the action. sí no

1. Cuando está de vacaciones, normalmente tu instructor(a) tiene más tiempo.
2. Cuando compra zapatos deportivos nuevos, normalmente tu instructor(a) hace más ejercicio.

Similar to the previous activity, you will now hear sentences. Indicate whether, in your opinion, your instructor regularly does the following things. Keep in mind that the form of the verbs following *cuando* in 13-24 encodes the habitualness of the action.

1. Cuando llama por teléfono, normalmente tu instructor(a) charla con su hermana.
2. Cuando cocina algo elegante, normalmente tu instructor(a) tiene invitados.

Indicate whether, in your opinion, it is true (verdad) or false (falso) that a typical university student will do the following things. Keep in mind that the form of the verbs following *cuando* in 1- 6 encodes the uncertainty of the time.

1. _____ Cuando empiece el verano, va a ir a la Florida.
2. _____ Cuando se levante por la mañana, va a ir al gimnasio.

Similar to the previous activity, you will now hear sentences. Indicate whether, in your opinion, it is true (verdad) or false (falso) that a typical university student will do the following things. Keep in mind that the form of the verbs following *cuando* in 1- 6 encodes the uncertainty of the time.

1. _____ Cuando regrese a la casa de sus padres, va a recibir muchos regalos.
2. _____ Cuando estudie, va a sacar buenas notas (grades).

Indicate whether, in your opinion, it is true (verdad) or false (falso) that a typical university student regularly does the following things. Keep in mind that the form of the verbs following *cuando* in 1-6 encodes the habitualness of the action.

1. _____ Cuando empieza el verano, normalmente va a la Florida.
2. _____ Cuando se levanta por la mañana, generalmente va al gimnasio.

Similar to the previous activity, you will now hear sentences. Indicate whether, in your opinion, it is true (verdad) or false (falso) that a typical university student regularly does the following things. Keep in mind that the form of the verbs following *cuando* in 1-6 encodes the habitualness of the action.

1. _____ Cuando regresa a la casa de sus padres, normalmente recibe muchos regalos.
2. _____ Cuando estudia, generalmente saca buenas notas (grades).

APPENDIX D**Examples of Assessment Tasks**

Assessment I - You will hear the first part of a sentence. Listen carefully and select the appropriate phrase that correctly completes each sentence.

1. Cuando Carmen anda por la ciudad...
 - a. se pierde (gets lost).
 - b. va a perderse (gets lost).
2. Cuando Pilar termine su lectura...
 - a. da un paseo
 - b. va a dar un paseo.

Assessment II - You will hear the beginning of a sentence. Please write your ending to the sentence using the verb provided.

1. Mi madre normalmente le ayuda cuando...Jorge (regresar)_____.
2. Juan va a sacar el perro cuando...él (escuchar)_____.

Assessment III - Select the appropriate phrase that correctly completes each sentence.

1. Cuando el niño esté enfermo ...
 - a. llora.
 - b. va a llorar.
2. Cuando Clara lleve su pesada mochila (*backpack*)...
 - a. la espalda (back) le duele.
 - b. la espalda (back) va a dolerle.

Assessment IV - Conjugate the verb in parenthesis to correctly complete the sentence.

1. David va a ganar mucho dinero cuando _____ (trabajar) allí.
2. Cuando Carla _____ (tomar) mucho café, generalmente no puede dormir bien.

AUTHOR

Erin McNulty Diaz, Ph.D., Assistant Professor, Department of Spanish and Portuguese, Dickinson College.

Correlating Aptitude with Oral Proficiency: A Quantitative Study of DLAB and OPI Scores across Four Language Categories

ISABELLE POUPARD SANTIZO

Defense Language Institute Foreign Language Center

This quantitative study focuses on the relationship between foreign language learners' aptitude and proficiency test scores. Four groups of 136 beginning students received six months of Initial Acquisition Training (IAT) in four different language categories, according to the level of complexity for an English speaker: French (Category I), Indonesian (Category II), Tagalog (Category III), and Arabic (Category IV). This study attempts to identify variations within the correlation among those four languages by analyzing the relationships between aptitude and proficiency scores from each category. The data gathered consisted of students' Defense Language Aptitude Battery (DLAB) and Oral Proficiency Interview (OPI) test scores. Results showed that the correlation between DLAB and OPI was moderate amongst French and Tagalog languages, the correlation with respect to Indonesian was almost nonexistent, and null for Arabic. The complexity of the foreign language categorization process, tests' reliability issues, and a fixed course length regardless of language category are some factors that could have contributed to the calculation of different correlation coefficients. Further studies based on experimental research are recommended to help substantiate causality to the findings derived from this study.

Keywords: *language aptitude, language proficiency, language categories, Defense Language Aptitude Battery (DLAB), Oral Proficiency Interview (OPI), correlation between DLAB and OPI*

INTRODUCTION

The vast majority of military linguists receive their initial language education at the Defense Language Institute Foreign Language Center (DLIFLC) in Monterey, California, where on any given day approximately 800 faculty members teach 22 languages to 3300 students (Lett, 2005). During the program of instruction, students develop foreign language and culture proficiency necessary for their mission. The Defense Language Aptitude Battery (DLAB) test, administered prior to training, helps determine into which foreign language students will be enrolled. In addition to its function as a placement test, DLAB scores are defined as a measure to indicate likelihood of success in learning languages *structurally close* to English (Child, 1998a).

The DLAB is a 90-minute, 119-item auditory multiple-choice test that requires examinees to learn and use an artificial language. This aptitude test measures the test-taker's strategies to extract and organize semantic, syntactic, and phonetic structure of language (Silva & White, 1993). The test was initially validated in 1971 on a sample of 879 graduates from 12 language courses (Petersen & Al-Haik, 1976). Since then, several studies have corroborated the test's validity. Recently, Planchon and Ellis (2014) conducted a study using the DLAB to establish language-learning aptitude based on bilingualism amongst 142 officers. Their study concluded that the DLAB is a reliable predictor of an individual's ability to learn a foreign language and is designed to help an organization determine where best to place its resources (Planchon & Ellis, 2014).

“Since its inception the use of the DLAB has evolved from predicting success at DLI [Defense Language Institute] to helping the institution sort learners into specific language categories applicable for the range of DLAB scores attained” (Wagener, 2015, p. 195). Language categorization is the division of languages into groups according to their level of complexity (compared to English) (Wagener, 2015). In other words, high DLAB scores indicate a strong likelihood of a student successfully learning a more complex foreign language, such as Arabic. There have been, over the years, several attempts to categorize languages in terms of their presumed difficulty; which is to say, how hard they are to learn for native speakers of English (Child, 1998b). “Category I languages are considered the easiest to learn for native speakers of English, followed by Category II, III, and then IV” (Wagener, 2015, p. 33). DLI currently establishes a 36 week course for category I and II languages (e.g., Spanish, French and Indonesian), 48 weeks to category III (e.g., Tagalog), and 64 weeks for category IV (e.g., Arabic), allowing more time for students to learn more complex languages (DLIFLC, 2016).

After completing the language training, military linguists take proficiency tests that assess their listening, reading, and speaking skills. Although the test of speaking has a long history, the direct testing of second language oral proficiency did not become commonplace until the 1980s. Oral

interviews developed by the Foreign Service Institute (FSI) and associated U.S. Government agencies have been long hailed as valid direct tests of speaking ability (Fulcher, 1996). The Oral Proficiency Interview (OPI) addresses several oral skills simultaneously, measuring from a global perspective, rather than from the point of view of the absence or presence, control or semi-control, of any given linguistic point (Parry & Child, 1990). The OPI is defined as an integrated approach that makes it possible for test-users to make a variety of inferences about the capacity for language use that test-takers have, or about what they can or cannot do (Bachman, 2002).

The effort of the American Council on the Teaching of Foreign Languages (ACTFL) to improve foreign language learning in the U.S. has increased the level of awareness of language educators about performance tests (Salaberry, 2000). The ACTFL OPI is a 20-30 minute one-on-one interview between a certified OPI tester and an examinee (ACTFL, 2016). “ACTFL currently uses the Oral Proficiency Interview (OPI) test to assess candidates in 48 languages for U.S. Government agencies and contractors and is constantly increasing the number of languages in which OPIs are available” (Kennedy & Stansfield, 2010, p. 4). “The ACTFL [OPI] guidelines are based on the Interagency Language Roundtable (ILR) descriptions of language proficiency for use in governmental and military organizations and have been modified for use in academia and industry. The ACTFL rating scale describes four major levels of language proficiency—Superior, Advanced, Intermediate, and Novice—that are delineated according to a hierarchy of global tasks related to functional language ability” (Surface & Dierdorff, 2003, p. 508).

There are three interconnected components that play a major role in a military linguist’s language training: the DLAB test, the foreign language categorization process, and the OPI test. The purpose of this study is to investigate the relationship between students’ language aptitude and language proficiency pertaining to different categories. First, the study will identify the correlation between the DLAB and the OPI test scores in four languages: French (Category I), Indonesian (Category II), Tagalog (Category III), and Arabic (Category IV). Second, correlation coefficient in each language will be compared to detect variations amongst the four languages. Finally, possible factors that contribute to the variations are provided, based on literature review.

RESEARCH QUESTION AND HYPOTHESES

When initially compared and informally discussed amongst faculty, students and staff, DLAB (aptitude) and OPI (oral proficiency) test scores obtained from a foreign language beginner IAT (Initial Acquisition Training) program seemed to show discrepancies for students enrolled in Arabic. Indeed, based on these initial results and faculty and student perception, high aptitude test scores did not seem to consistently translate into high proficiency test scores for students who were learning to speak Arabic during a six-month course. To validate non-empirical data that include forms of introspection, vicarious

experiences, and other people's analysis of events, this study attempts to answer the following research question: After a six month IAT course, is the correlation coefficient of aptitude (DLAB) and proficiency (OPI) test scores less significant for Arabic than it is for French, Indonesian, and Tagalog?

This quantitative study will collect and analyze the data (DLAB and OPI scores) focusing on the following hypotheses:

H0: There is absence of correlation amongst variables (DLAB and OPI scores) within each of the following languages: Arabic, French, Tagalog and Indonesian.

H1: The correlation coefficient of aptitude (DLAB) and proficiency (OPI) test scores is lower for Arabic when compared to the other three languages—French, Indonesian, and Tagalog.

By addressing H1, this study attempts to identify variations within the correlations among four language categories and provide potential explanations to address the extent to which correlations in Arabic language might differ from correlations found in the other three languages (French, Indonesian and Tagalog).

LITERATURE REVIEW

Criticism Towards the Oral Proficiency Interview

Empirical studies have examined the OPI's reliability and validity. The findings from a qualitative study focusing on the guidelines, construct, testers, and design of the OPI test concluded that the OPI does measure distinct skills and is construct valid (Lazaraton, 2002). Another study examined 5,881 interviews in 19 languages from four language categories and reported that "educators and employers who use the ACTFL OPI can expect reliable results and use the scores generated from the process with increased confidence" (Surface & Dierdorff, 2003, p. 519).

Even though the validity of the OPI test is widely accepted, how pertinent aspects of the validity and reliability are to be investigated and established is still problematic (Lazaraton, 2002). There have been criticisms regarding oral interviews; dilemmas faced by any attempt to measure language proficiency remain although a greater understanding of the nature of language—how it is used and how it is learned—has enhanced our insights into the constructs that we measure as language testers (Alderson & Banerjee, 2002). Today, an impediment to research for language testers, and a source of uncertainty for OPI users, is still the lack of a clearly defined speaking proficiency construct underlying the OPI and its attendant rating procedure (Amoroso, 2015).

Previous research established that government agencies had high inter-rater reliability for proficiency ratings within a given agency, but that the

standards, or their interpretations, were not always the same at every agency (Ehrman, 1998). Almost 20 years later, some of the criticisms towards the OPI continue to be the standardization process involved in foreign language testing. “The open-ended nature of the ACTFL OPI makes the contextual appropriateness of examinee speech especially difficult to evaluate automatically: no two tests are the same, so one cannot write code that can reliably tell the computer what to look for” (Amoroso, 2015, p. 165).

Limitations of the DLAB

Ascertaining limitations of aptitude tests, Ehrman (1990) suggested that the aptitude test tended to treat learning aptitude as an internal characteristic determined by an individual difference or trait. It would have made more sense not to think of testing as a measurement of an aptitude trait, but as prediction of language learning success in a more general way. Lowe (1998) conducted a study on various aptitude tests to address their construct validity. “The fact that past aptitude test designers have drawn on rather divergent predictors to determine aptitude for learning another language suggests either that there is no agreement on what constitutes the construct *language aptitude*, nor that there exist numerous possible predictors that could serve as components in determining the construct” (Lowe, 1998, p. 22).

Restrictions of the DLAB test have also pointed out a discrepancy between the skills that are tested for ability versus those tested for performance (Lowe, 1998). Lowe (1998) also concluded it to be impossible to predict language learning ability in a particular skill modality unless the aptitude test contained tasks targeting that modality. The nature of the DLAB is considered paradoxical as it only targets the reading skill but predicts success in all four skills, which suggests the format of the DLAB should be revisited. “The import for aptitude test design is that we should have a separate listening, reading, speaking, or writing component or even separate tests, if we desire to make statements about a given skill modality” (Lowe, 1998, p. 32). A recent study focusing on Chinese language learners’ aptitude determined that “it may profit the field to take aptitude research to another level and investigate aptitude for dynamic complexity” (Winke, 2013, p. 110).

Rethinking the Foreign Language Categorization

Child (1998b) suggested that despite the great benefits to the field of foreign language, the language categorization process does not specify what features of which languages can be expected to cause trouble for learners and which are similar to, or not very different from, comparable English features. “The entire language aptitude enterprise could falter in the absence of a comprehensive overview of similarities and differences among the major languages of the world... To lend greater precision to a *global assessment* system it is necessary to determine which major linguistic features of the so-

called *hard* languages make learning problematic for English speakers and which lend themselves to (relatively) easy transfer into English” (Child, 1998b, p. 16).

For years, DLIFLC has advised the services to consider both the language difficulty categories and individuals’ DLAB scores in selecting potential language students and assigning them to specific languages (Lett & O’Mara, 1990). However, recent criticisms continue to surface regarding the views on rating language complexity. Addressing the DLAB predictability of proficiency at lower levels and examining how DLAB aligns with DOD language categories, Wagener (2015) suggested “languages could be categorized by the aptitude components that best predict learner success in the language” (p. 212).

Correlating DLAB and OPI Scores

Different language categories have substantiated a correlation between the DLAB and the OPI results. “Earlier language aptitude tests attempted to define and operationalize each designer’s concept of the construct” (Lowe, 1998, p. 25). Empirical findings have identified a positive coefficient correlation between aptitude (DLAB) and proficiency (OPI) test scores. Petersen and Al-Haik (1976) established validity through prediction of grades. A later study conducted by Silva and White (1993), which examined the validity of the DLAB scores of 5,673 military students, concluded that the DLAB contributed significant incremental validity beyond general aptitude (Silva & White, 1993).

Notwithstanding the validated correlation previously established by research between the DLAB and the OPI scores, “an approximate .50 correlation between aptitude test scores and exit proficiencies suggests that one might be able to do so more fully” (Lowe, 1998, p. 25). Indeed, Lowe (1998) brought to light many questions that the aptitude test design has failed to answer and queried whether it would be possible to return to ground zero and design tests that could address them. A previous study involving another aptitude test, the Modern Language Aptitude Test (MLAT), concluded that correlations between aptitude and proficiency were stronger for Category I languages than for Category II, III, and IV languages (Ehrman, 1998).

More recently, a study involving over 9,000 students indicated that as performance standards change, it would be important to revisit how the DLAB and other student characteristics predict success (Schmitz, Stoloff, Wolfanger, & Sayala, 2009). Another investigation identified a small to moderate magnitude between the two tests, indicating that within a given language the differences in OPI ratings are largely due to factors other than language learning aptitude (Watson, Harman, & Surface, 2012). Similarly, a study conducted by Yeon (2014) revealed that successful students have several traits in common, including high aptitude test scores. Whereas a correlation between the DLAB and the OPI scores was verified in Category III and IV languages, the DLAB

and the OPI scores did not show any significant relation when all languages were combined (Yeon, 2014).

METHODS

Sample

All learners of the current study stemmed from the same learning context. The data are originated from a six-month foreign language program that required beginner military students to take the DLAB as a course pre-requisite and the ACTFL OPI following their training. Scores were compiled from several foreign language courses conducted by a military training school in the Southern U.S., of the same length, taught from 2011 through 2016. Students were enrolled in four different languages: French (Category I), Indonesian (Category II), Tagalog (Category III), and Arabic (Category IV). There was no requirement for a minimum DLAB score for students to enroll in the program, as compared to prerequisites established in other IAT guidelines, because students were required to complete the course as part of their overall military training. Those with higher DLAB scores were placed in more complex languages (in this case, Arabic or Tagalog). Graduation requirement for this program was set at Level 1 on the Interagency Language Roundtable (ILR) scale in an Oral Proficiency Interview (OPI) conducted by the ACTFL.

The gathered data consist of 272 scores from 136 pairs of aptitude (DLAB) and proficiency (ACTFL OPI) test scores (see appendix). On average, Arabic linguists constitute half of the student enrollment, with French, Indonesian, and Tagalog making up the other half.

A proportionally stratified sampling technique was used—the sample collected followed the same percentage of students enrolled in each language. The data was arbitrarily selected from a database by computer, using random number generator selection that associated each student with a number and a letter, maintaining student anonymity. Students enrolled in Arabic language were given the language designator AD, followed by a number (e.g., AD1, AD2, AD3, etc.). Students enrolled in French, Indonesian, and Tagalog had the language designator FR (FR1, FR2, FR3, etc.), JN (JN1, JN2, JN3, etc.), and TA (TA1, TA2, TA3, etc.). The data was distributed according to the population's percentage as follows: 136 DLAB scores, 68 OPI test scores of Arabic, 23 OPI test scores of French, 23 OPI test scores of Indonesian, and 22 OPI test scores of Tagalog. Thus, the total number, including aptitude and proficiency test scores, is 272.

Analysis

When determining the sample size, the Cohen's Power Tables were used and the highest power level (0.99 or 99%) was chosen, which diminishes the margin of error. In this study, there was a 1% chance that significant

differences between the correlation of the DLAB and the OPI scores could be missed.

The mean and the standard deviation of the DLAB and the OPI scores were calculated. The DLAB scores were entered as a scale measurement, which ranges from 0-164. The OPI scores gathered in this study range from 1 to 3 on the ILR skill level, which establishes Level 1 as elementary proficiency, Level 1+ as elementary proficiency plus, Level 2 as limited working proficiency, Level 2+ as limited working proficiency plus, and Level 3 as general professional proficiency (Interagency Language Roundtable, 2016). The OPI scores were entered as a scale measurement ranging from 1-5 as follows: 1 as elementary proficiency, 2 as elementary proficiency plus, 3 as limited working proficiency, 4 as limited working proficiency, and 5 as general professional proficiency

A high standard deviation of the DLAB and the OPI scores shows that scores vary consistently within each item calculated, meaning, for instance, that language aptitudes are not homogenous. It is important to see if a high standard deviation of the DLAB scores translates into a high standard deviation of OPI scores.

The study relied on Statistical Package for the Social Science (SPSS) for data processing. Cronbach's Alpha was used to identify if there is a variation of correlation between the items that measure the aptitude and the proficiency scores across four languages. Identifying a strong relationship between the test scores not only provided instrument reliability, but also helped us address the research question.

RESULTS OF HYPOTHESIS TESTING

Table 1 illustrates the aptitude scores (DLAB) and Table 2 the proficiency scores (OPI) for all students combined. The minimum DLAB score in this study is 64, the maximum 135, and the mean 94.23, with a standard deviation of 17.8. The maximum OPI score is 4 (2+ ILR), the minimum 1 (1 ILR), and the mean 1.87, with a standard deviation of 0.71. The results indicate that both scores are not too widespread from the mean (94.23 for the DLAB and 1.87 for the OPI). The smaller standard deviation indicate that the mean is representative of all individuals in the group.

Table 1
DLAB Scores

	<i>N</i>	<i>Minimum</i>	<i>Maximum</i>	<i>Mean</i>	<i>Std. Deviation</i>
<i>Aptitude Scores</i>	136	64.00	135.00	94.2353	17.81062
<i>Valid N (listwise)</i>	136	Valid N (listwise)			

Table 2
OPI Scores

	<i>N</i>	<i>Minimum</i>	<i>Maximum</i>	<i>Mean</i>	<i>Std. Deviation</i>
<i>Proficiency Scores</i>	136	1	4	1.87	.718
<i>Valid N (listwise)</i>	136				

A two-tailed test was used to detect additional relationships between the test scores. In other words, it is important to see if the DLAB scores vary according to the OPI scores, or if the DLAB scores fluctuate in the opposite direction of the OPI scores. The OPI scores did not go up comparably to the DLAB scores. Table 3 shows the data of all participants ($n = 136$). The overall correlation is .023, which indicates that there is no correlation between the DLAB and the OPI scores when all four languages (Arabic, French, Indonesian, and Tagalog) are combined. We observe that, by aggregating the scores, the null hypothesis (H_0) cannot be rejected. Sig. (2-tailed) value is 0.788. Thus, when all languages were combined, the findings, in addition to indicating a nonexistent correlation coefficient, also verified no statistical power.

Table 3
Overall Correlations

		<i>Aptitude Scores</i>	<i>Proficiency Scores</i>
<i>Aptitude Scores</i>	Pearson Correlation	1	.023
	Sig. (2-tailed)		.788
	<i>N</i>	136	136
<i>Proficiency Scores</i>	Pearson Correlation	.023	1
	Sig. (2-tailed)	.788	
	<i>N</i>	136	136

$P < .05$

When students are separated into four groups according to language categories, the level of the correlation varies. As shown in Table 4, there is a .204 correlation coefficient for Indonesian students. This low correlation coefficient reveals almost no relationship between the DLAB and the OPI scores among Indonesian students. Sig. (2-tailed) of .350 indicates that there is no statistically significant difference. Similarly, when the scores are aggregated, the null hypothesis cannot be rejected when treating the scores of Indonesian students.

Table 4
Indonesian Correlations

		<i>DLAB</i>	<i>OPI</i>
	Pearson Correlation	1	.204
<i>DLAB</i>	Sig. (2-tailed)		.350
	<i>N</i>	23	23
	Pearson Correlation	.204	1
<i>OPI</i>	Sig. (2-tailed)	.350	
	<i>N</i>	23	23

$P < .05$

Table 5 illustrates the aptitude and proficiency scores for Tagalog students with a correlation coefficient of .542. This correlation coefficient indicates a moderate relationship between the DLAB and OPI scores after a six-month IAT. This is a moderate correlation with statistically significant because f (significance) value is very low, Sig. (2-tailed) of .009. Contrary to Indonesian scores, the null hypothesis can be rejected because a correlation exists for Tagalog.

Table 5
Tagalog Correlations

		<i>DLAB</i>	<i>OPI</i>
	Pearson Correlation	1	.542
<i>DLAB</i>	Sig. (2-tailed)		.009
	<i>N</i>	22	22
	Pearson Correlation	.542	1
<i>OPI</i>	Sig. (2-tailed)	.009	
	<i>N</i>	22	22

$P < .05$

Table 6 shows the relationship between the test scores among 23 French students. SPSS calculated a .529 correlation coefficient, indicating a moderate relationship between the DLAB and the OPI scores. Similarly to Tagalog, Sig. (2-tailed) of .009 indicates statistical significance for French scores.

Table 6
French Correlations

		<i>DLAB</i>	<i>OPI</i>
	Pearson Correlation	1	.529
<i>DLAB</i>	Sig. (2-tailed)		.009
	<i>N</i>	23	23
	Pearson Correlation	.529	1
<i>OPI</i>	Sig. (2-tailed)	.009	
	<i>N</i>	23	23

P < .05

As shown in Table 7, results derived from 68 Arabic students indicate the lowest correlation coefficient at .144, revealing a close to 0 relationship between the DLAB and the OPI scores. Sig. (2-tailed) of .241 shows that there is no statistically significant difference. Similar to Indonesian and to the analysis with aggregated scores, the null hypothesis cannot be rejected when treating the scores of Arabic students.

Table 7
Arabic Correlations

		<i>DLAB</i>	<i>OPI</i>
	Pearson Correlation	1	.144
<i>DLAB</i>	Sig. (2-tailed)		.241
	<i>N</i>	68	68
	Pearson Correlation	.144	1
<i>OPI</i>	Sig. (2-tailed)	.241	
	<i>N</i>	68	68

P < .05

All coefficients calculated in this study show a positive correlation, suggesting there are no inverse relationships between the DLAB and the OPI scores. All languages combined, the .023 coefficient indicates the absence of correlation between the DLAB and the OPI scores. The absence of correlation for all languages combined as well as for Arabic in isolation and a very low relationship between the DLAB and the OPI scores for Indonesian are examined in the discussion part of this paper.

When all languages are combined the .023 coefficient indicates there is an absence of correlation between the DLAB and the OPI scores, thus the null hypothesis cannot be rejected. The statistical intricacy of this quantitative study resides in the fact that identifying four correlation coefficients constitutes the first step; comparing the subsequent four correlation coefficients is the second and most important step because it directly addresses the hypotheses. In this

context, a null hypothesis (H0) would translate into the absence of deviations within correlations. The results indicate that when separated into four language groups, two out of four coefficients show variations amongst them. In other words, the null cannot be rejected with respect to Tagalog and French coefficients.

However, the overall variations confirm the hypothesis (H1). Because a difference is found among Arabic (no correlation), Indonesian (very weak correlation), and French and Tagalog (moderate correlation), the correlation coefficient of aptitude (DLAB) and proficiency (OPI) test scores is lower for Arabic when compared to the other three languages—French, Indonesian, and Tagalog (H1).

When the correlation coefficients of the four language groups are calculated and compared, two out of four coefficients show variations amongst them. With respect to Tagalog and French, there is a moderate correlation between the DLAB and the OPI scores, which presents statistical significance ($f=.009$). Additionally, a low p value decreases the margin of error for these two languages limiting to less than 1% probability of the correlation happening by chance alone. Thus, in the case of French and Tagalog, the moderate correlation between DLAB and OPI has strong statistical power. For Arabic and Indonesian, the correlation coefficients are very low (.204 for Indonesian) to null (.144 for Arabic), meaning there is a very weak correlation between the DLAB and the OPI scores for students enrolled in these two languages. A high p value for these two languages decrease the statistical power of the correlation between the test scores, indicating that for Indonesian there is a 35% chance that other factors have influenced the test scores, whereas the probability of the correlation happening by chance alone is high at 24% for Arabic.

DISCUSSION

The research question investigated whether the correlation coefficient of the aptitude (DLAB) and the proficiency (OPI) test scores are less significant for Arabic than for French, Indonesian, and Tagalog after a six-month IAT course—the answer is yes. The hypothesis (H1) in this study predicts a lower coefficient correlation for Arabic when compared to French, Indonesian, and Tagalog. The results have not only confirmed H1 but also have (1) suggested a null correlation of the test scores for Arabic; (2) revealed a very weak correlation for Indonesian; and (3) indicated only a moderate correlation for French and Tagalog.

Considering that the correlation coefficient for Arabic is null and that Arabic students represent half of the student population in the language training program, we can explain why the overall correlation for the 136 students was 0 (.023 meaning there is no relationship between the DLAB and the OPI scores). A significant finding in this study is that levels of correlation vary drastically once languages are separated, suggesting that the correlation varies in each language. A null or a very low correlation coefficient between the DLAB and

the OPI scores can be problematic because it questions the accurate prediction of foreign language learning success. A correlation coefficient ranging from .6 to .8 would have enabled us to expect that a strong relationship existed between the two test scores and to believe that the DLAB is accurate in predicting student oral proficiency success in the context of a six-month beginner course.

As suggested in the literature review section of this study, there are limitations in the language categorization and the aptitude and proficiency testing processes, which might partially explain the absence of correlation for Arabic and a very low relationship for Indonesian. Lowe (1998) verified a correlation between DLAB scores and student exit proficiencies, but also identified varying coefficients according to individual skill modalities and test design, leading to his observation—“this variance again raises the question, what is the focus of the test?” (Lowe, 1998, p. 46). Wagener (2015) investigated whether predictive measures can differentiate levels of language proficiency for learners across languages, language categories, and learning contexts. He concluded that predictive models of proficiency are not consistent within a language category, nor are they consistent across language category boundaries. Additionally, his research provided “evidence that predictor profiles of language learning success vary across individual foreign languages” (p. 204).

As stated earlier in this paper, DLAB scores help predetermine in which foreign language a student should be enrolled. The absence of a correlation coefficient for the Arabic language in this study could be explained by the fact that Arabic, along with three other languages (Chinese, Japanese, and Korean), is defined as a Category IV language—one of the most difficult languages for native English speakers to learn. These languages “share the difficulty of different writing systems and of non-Western European culture, but from that point on there are more divergences than commonalities [between these languages]” (Child, 1998b, p. 27). Thus, although paradoxical, the main criteria for clustering Arabic as a category IV language is based on one skill (writing); this might seem irrelevant when assessing another skill (speaking), as in the case of the OPI test. Wagener (2015) recommended that further research be conducted to determine if a re-categorization of languages would better align predictor success within the language category structure. “Languages could be categorized by the aptitude components that best predict learner success in the language” (Wagener, 2015, p. 210).

Another parallel that can be drawn between the very low correlation coefficients identified in this study and previous research is the categorization of Indonesian, a Category II language. This language is considered among those easier for Americans to learn. In fact, Indonesian “might be regarded as a Category 1 language if one wishes to achieve any level between 0 and 2+, but proves to be a Category 2 language if one wants to go beyond 2+ into 3, 3+, 4, 4+, or 5! Can an aptitude test predict success for languages of shifting degrees of difficulty such as Indonesian?” (Lowe, 1998, p. 30). “The division of languages into difficulty groups aids in planning training, but it clusters together languages whose common features may cause Americans difficulties in learning, yet

whose nature can differ radically in structure and thought patterns from language to language” (Child, 1998b, p. 25).

Null, low, and moderate levels of correlation found in this study could also be a result of the skills and levels pertinent to each language program. A robust study focusing on advanced Chinese language identified a variation of how aptitude impacts skills differently. The research revealed that aptitude, strategy use, and motivation had about the same impact on learning but varied in how well they predicted the individual skills of listening, reading, and speaking (Winke, 2013). The findings suggest that “high aptitude, high motivation, and good strategy use may be significantly advantageous conditions for attaining advanced proficiency, but when instruction is task-based and grounded in social interaction, minute distinctions in advanced proficiency may depend more on unmeasurable and unsystematic factors external to the model” (Winke, 2013, p. 121). Identifying effects of aptitude levels on different skills in an initial acquisition context would definitely expand the reach of the present correlational study.

Another factor that might have contributed to the present results can be attributed to testing reliability issues, for the ACTFL OPI in particular where “some operational measures are still considered poor predictors of OPI ratings (grammatical accuracy, abstractness vs. concreteness, and cohesion), while others are quite robust (vocabulary and fluency), while yet others are not measured at all (pronunciation and register shift) due to computational constraints” (Amoroso, 2015, p. 165). Surface and Dierdorff (2003) identified discrepancies when examining OPI scores from Arabic and Italian students in particular. Their study came to the conclusion that “both [languages] had slightly lower levels of rater consistency” (Surface & Dierdorff, 2003, p. 514).

Finally, the course length most likely had an impact on the OPI scores used in this correlational study. In a foreign language program for military linguists, each language category is typically given a different length of instruction according to its level of complexity. Indeed, the more easily transferable a foreign language is to English speakers, the less time it takes to learn it. Due to the unique nature of this Initial Acquisition Training, students were given the same amount of time to learn a language, regardless of its complexity, as this foreign language course is integrated into other military training required by Marines. Thus, students of Arabic (a Category IV language) were expected to meet the same graduation requirements as students enrolled in French (a Category I language) in the same amount of time.

It is important to note that this study does not attempt to compare proficiency levels across foreign languages. Instead, its goal is to identify variations and analyze the relationships between aptitude and proficiency scores in each language category. Given that the acquisition of higher category languages requires more time, it seems contradictory that correlation is not detected in both Arabic (Category IV) and Indonesian (Category II), yet is found in Tagalog (Category III) and French (Category I), showing an inconsistency in the relationship between scores and language categories; hence Lowe’s question

is still pertinent today: “can one predict who will be most successful, not just learning a language, but learning one of a specific category?” (Lowe, 1998, p. 27).

A recent report on DLAB as a predictor of foreign language learning by Watson, Harman, and Surface (2012), involving more than 1,800 students enrolled in the Initial Acquisition Program varying in length according to language complexity, found similar discrepancies as those found in this study. “For instance, trainees of all language aptitude levels were more likely to attain ILR 2 in Modern Standard Arabic (Cat IV) than Russian (Cat III). Similarly, more trainees attained ILR 2 in Indonesian (Cat II) compared to Spanish (Cat I) and French (Cat I)” (p. 9). Whereas languages can be grouped depending on how much time it takes to learn them—the current basis of grouping languages by difficulty—“languages can also be grouped depending on the kinds of difficulties they involve” (Child, 1998b, p. 27).

RECOMMENDATIONS

Although atypical, other foreign language programs also conduct training following a set course length regardless of the language category, particularly in the Special Force (SF) community. Students are usually enrolled in shorter courses with the purpose of acquiring the necessary language capabilities to achieve a level of proficiency distinct from the vast majority of linguists. In fact, “SF personnel are required to learn and maintain a language in order to enhance their core mission activity. They are considered language-enabled personnel, not linguists, since language is not their core mission” (Ellington, Surface, Blume, & Wilson, 2015, p. 41). Nonetheless, despite the change in the course length, the training is not military specific because it focuses on general language proficiency. To calculate the impact of course lengths on correlation coefficients of aptitude and proficiency, further correlational studies are recommended. As a comparison to the data found in this study, data can be collected from students whose course length varies according to language complexity.

Proceeding to a deeper understanding of why the relationship between aptitude and oral proficiency is close to zero for Arabic and Indonesian would allow a possible shift in the approach of how the DLAB is administered as a placement test to achieve foreign language learning success at its early stages. “One place to start is to investigate aptitude constructs individually and within a larger picture of cognitive, cognitively oriented, and affective variables and situated within a particular language learning context” (Winke, 2013, p. 110). The results of this study suggest that if a moderate correlation exists in French and Tagalog, none in Arabic, and very small in Indonesian, students with a high DLAB score might achieve higher OPI scores when placed in French and Tagalog, but not in Arabic and Indonesian. “This last point demonstrates the possibility of better alignment of learners with specific abilities into languages that are more demanding of those abilities” (Wagener, 2015, p. 210). Finally,

according to the findings of this study, it is detrimental to a study's reliability to aggregate all languages together when analyzing the data.

As the ability to better predict the OPI results based on DLAB scores across language categories would increase foreign language learning success, further studies are recommended to identify possible causality that can be derived from the present findings. The correlations identified in this study indicated the strength of relationship between two variables: DLAB and OPI scores. A regression study could pay closer attention to the inter-reliability of the three instruments of aptitude (DLAB), proficiency (OPI), and language categorization based on its difficulty. For instance, can the weak correlation between the DLAB and the OPI scores for Indonesian be explained by the intricacies of language categorization? Does the absence of correlation in Arabic reveal a need to review how students are placed in different language categories based on their aptitude scores? A carefully planned experimental study is recommended to establish causality through the manipulation of one of the variables (DLAB or OPI scores) while controlling all other possible extraneous variables that might have contributed to the oscillation of the scores (language categories). More research would generate stronger substantiating evidence of DLAB and OPI score variations across language categories while considering the limitations of language categorization.

The implications of the current findings point towards a review of using the DLAB as predictor of foreign language oral proficiency success in the context of a six-month beginner course. Currently, "several DOD language training programs do not see the value of the DLAB at predicting success in lower proficiency learners and are using other aptitude measures for program selection decisions" (Wagener, 2015, p. 41). Efforts to improve the DLAB predictability are now in place with "the DLAB2 project [which] will develop a new version of the DLAB that is based on advances in cognitive science, personality and trait psychology, and foreign language education" (DLAB, 2016).

LIMITATIONS AND FUTURE DIRECTIONS

In terms of statistics, the present study is affected by a methodological limitation with regards to the participants from whom the scores were collected. The Cohen's statistical power table was used as one of the quantitative methodology guidelines for the present study when establishing a reliable sample size—there should have been 136 students per language category to reach the statistically recommended number. In fact, researchers would agree that a larger number of scores would have been preferable for validity purposes. Another point to consider is that not all correlations are reflected by a straight line showing a linear correlation between variables (Salkind, 2008). In this correlational study, the relationship is not linear because OPI scores are not considered as continuous rating. Indeed, the interval between a Level 1 and 2 to

a Level 2 and 3 are not the same. Thus, an exponential growth in proficiency measured by the ILR scales could not be captured.

Future programs may consider incorporating additional factors when measuring student aptitude. The implementation of a subtest developed at the language program level, for instance, could integrate other skills to match the ones on which students will be tested at the end of their training. Transferability of the current DLAB test from reading to listening and speaking could present a challenge, especially if each subtest's content varies according to the purpose of each language program (familiarization, initial acquisition, sustainment, and enhancement). It seems plausible to assert that because of the complexity involved in foreign language aptitude, the DLAB cannot be used as a one-size-fits-all. For instance, a separate DLAB can be designed for initial acquisition students, focusing on the oral proficiency skill. An integrated and comprehensive re-adaptation of the DLAB to mold its format and content by including skills, levels, and foreign language considerations would greatly benefit the IAT community.

Generalization of the findings to other learning contexts and student population is not recommended. The research is limited by the data from a six-month program, as opposed to other programs that establish different lengths of training. Another unique nature of the program where data were collected is that it does not allow students to drop or to be dropped from their language training. Regardless of students' abilities and faculty recommendation, Marines are required to continue the course and are expected to reach Level 1 (ILR) in the ACTFL OPI test by graduation. Additionally, contrary to other IATs, this program does not require a minimum DLAB score in order to place a student in language instruction. Moreover, graduation requirement is limited to ILR Level 1 although many exceed the requirement. Finally, this study does not involve proficiency in terms of other skills (reading and listening). Thus, the findings in this study may not be relevant to other types of assessments involving listening and reading such as the Defense Language Proficiency Test and the Diagnostic Assessment.

Nevertheless, the relevance of this study encompasses not only foreign language students and instructors but also program managers. Most importantly it suggests that closer attention be paid to how well and how much can be measured from the available resources when predicting foreign language learning success. Stakeholders have called for continuous improvement in foreign language training, from the selection process (student enrollment) to the training outcome (proficiency tests). It is valuable to revisit the available criteria—aptitude and proficiency tests—currently in place with respect to the foreign language categorization process.

REFERENCES

- Alderson, J. C., & Banerjee, J. (2002). State of the art review: Language testing and assessment (part two). *Language Teaching*, 35(2), 79-113. 10.1017/S0261444802001751
- American Council on the Teaching of Foreign Languages (2016). *ACTFL proficiency guidelines 2012*. Retrieved from <https://www.actfl.org/publications/guidelines-andmanuals/actfl-proficiency-guidelines-2012>
- Amoroso, L. W. (2015). *I don't know what it is, but I know it when I hear it: Speaking proficiency and the ACTFL OPI* (doctoral dissertation, Georgetown University). Retrieved from <https://repository.library.georgetown.edu/handle/10822/1029903>
- Bachman, L. F. (2002). Some reflections on task-based language performance assessment. *Language Testing*, 19, 453-476.
- Child, J. R. (1998a). Language skill levels, textual modes and the rating process. *Foreign Language Annals*, 3, 381-397.
- Child, J. R. (1998b). Language aptitude testing: Learners and applications. *Applied Language Learning*, 9(2), 1-10.
- Defense Language Aptitude Battery (2016). *Defense language aptitude battery 2 (DLAB2)*. Retrieved from <https://www.casl.umd.edu/projects/defense-language-aptitude-batter-2dlab2/>
- Defense Language Institute Foreign Language Center (2016). *Languages taught at DLIFLC and duration of courses*. Retrieved from <http://dliflc.edu/about/languages-at-dliflc/>
- Ehrman, M. (1990). The role of personality type in adult language learning: An ongoing investigation. In T. S. Parry, & C. W. Stansfield (Eds.), *Language aptitude reconsidered* (pp. 126-178). Englewood Cliffs, NJ: Prentice Hall Regents.
- Ehrman, M. (1998). The modern language aptitude test for predicting learning success and advising students. *Applied Language Learning*, 9(2), 31-70.
- Ellington, J. K., Surface, E. A., Blume, B. D., & Wilson, M. A. (2015). Foreign language training transfer: Individual and contextual predictors of skill maintenance and generalization. *Military Psychology*, 27(1): 36-51.
- Fulcher, G. (1996). Invalidating validity claims for the ACTFL oral rating scale. *System*, 24, 163-172.
- Interagency Language Roundtable (2016). *Interagency language roundtable (ILR) scale*. Retrieved from <http://www.govtilr.org/Skills/ILRscale1.htm>
- Kennedy, L., & Stansfield, W. (2010). The reading proficiency interview (RPI): A rapid response test development model for assessing reading proficiency on the ILR scale. *Applied Language Learning*, 20, (1/2), 1-16.
- Lazaraton, A. (2002). *A qualitative approach to the validation of oral language tests*. Cambridge: Cambridge University Press.

- Lett, J. (2005). Foreign language needs assessment in the US military. In M. H. Long (Ed.), *Second language needs analysis* (pp. 105–124). Cambridge: Cambridge University Press.
- Lett, J., & O'Mara, F. E. (1990). Predictors of success in an intensive language learning context: Correlates of language learning at the defense language institute foreign language center. In T. S. Parry, & C. W. Stansfield (Eds.), *Language aptitude reconsidered* (pp. 222–260). Englewood Cliffs, NJ: Prentice Hall Regents.
- Lowe, P. (1998). Zero-based language aptitude test design: Where's the focus for the test? *Applied Language Learning*, 9(2), 11–30.
- Parry, T., & Child, J. (1990). Preliminary investigation of the relationship between VORD, MLAT and language proficiency. In T.S. Parry, & C.W. Stansfield (Eds), *Language aptitude reconsidered* (pp. 30–66). Englewood Cliffs, NJ: Prentice Hall.
- Petersen, C., & Al-Haik, A. (1976). The development of the defense language aptitude battery (DLAB). *Educational and Psychological Measurement*, 36, 369–380.
- Planchon, A., & Ellis, E. (2014). A diplomatic advantage? The effects of bilingualism and formal language training on language aptitude amongst Australian diplomatic officers. *Language Awareness*, 23(3), 203–219.
- Salaberry, R. (2000). Revising the revised format of the ACTFL oral proficiency interview. *Language Testing*, 17(3), 289–310.
- Schmitz, E. J., Stoloff, P. H., Wolfanger, J. S., & Sayala, S. (2009). *Accession screening for language skills and abilities*. CNA Analysis and Solutions.
- Silva, J. M., & White, L. A. (1993). Relation of cognitive aptitudes to success in foreign language training. *Military Psychology*, 5, 79–93.
- Surface, E. A., & Dierdorff, E. C. (2003). Reliability and the ACTFL oral proficiency interview: Reporting indices of interrater consistency and agreement for 19 languages. *Foreign Language Annals*, 36, 507–519.
- Watson, A. M., Harman, R. P., & Surface, E. A. (2012). *Evaluating DLAB as a predictor of foreign language learning*. Raleigh, NC: SWA Consulting.
- Wagener, T. R. (2015). *The influences of aptitude, learning context, and language difficulty categorization on foreign language proficiency* (doctoral dissertation, University of Maryland). Retrieved from <http://drum.lib.umd.edu/handle/1903/17314>
- Winke, P. (2013). An investigation into second language aptitude for advanced Chinese learning. *Modern Language Journal*, 97, 109–130.
- Yeon, S. H. (2014). It works for me: from the high achievers' perspective. *Applied Language Learning*, 23–24, 75–86.

APPENDIX**STUDENT DLAB AND OPI SCORES**

STUDENT ID	DLAB SCORES	OPI SCORES
JN1	74	3
JN2	108	3
JN3	84	3
JN4	65	2
JN5	96	3
JN6	80	2
JN7	80	2
JN8	94	3
JN9	65	3
JN10	74	2
JN11	65	3
JN12	68	3
JN13	75	3
JN14	94	3
JN15	76	2
JN16	88	2
JN17	105	3
JN18	135	3
JN19	90	2
JN20	93	3
JN21	70	3
JN22	105	4
JN23	70	4
TA1	86	1
TA2	117	2
TA3	74	1
TA4	76	1
TA5	86	2
TA6	88	2
TA7	90	2

TA8	86	2
TA9	99	2
TA10	81	2
TA11	88	2
TA12	89	2
TA13	76	2
TA14	114	2
TA15	123	2
TA16	103	2
TA17	116	3
TA18	110	3
TA19	107	2
TA20	105	2
TA21	108	2
TA22	74	2
FR1	80	2
FR2	89	2
FR3	76	2
FR4	89	2
FR5	94	1
FR6	90	2
FR7	71	2
FR8	76	2
FR9	79	1
FR10	64	1
FR11	71	1
FR12	75	1
FR13	79	2
FR14	66	1
FR15	65	1
FR16	77	2
FR17	86	3
FR18	114	3
FR19	123	3

FR20	71	3
FR21	124	2
FR22	70	1
FR23	105	2
AD1	119	2
AD2	123	2
AD3	114	1
AD4	109	2
AD5	125	1
AD6	104	1
AD7	98	2
AD8	103	2
AD9	112	1
AD10	100	1
AD11	110	2
AD12	127	2
AD13	99	1
AD14	125	1
AD15	91	3
AD16	116	1
AD17	95	1
AD18	99	1
AD19	97	1
AD20	103	1
AD21	109	1
AD22	126	2
AD23	96	2
AD24	126	1
AD25	105	2
AD26	90	2
AD27	90	2
AD28	88	2
AD29	127	2
AD30	117	2

AD31	90	2
AD32	89	2
AD33	116	2
AD34	119	2
AD35	109	2
AD36	90	2
AD37	94	2
AD38	102	2
AD39	131	2
AD40	99	2
AD41	93	1
AD42	86	1
AD43	80	1
AD44	122	2
AD45	70	1
AD46	89	1
AD47	70	2
AD48	82	1
AD49	81	1
AD50	83	1
AD51	80	1
AD52	95	1
AD53	92	1
AD54	95	2
AD55	98	2
AD56	100	2
AD57	90	2
AD58	87	1
AD59	84	1
AD60	84	1
AD61	70	1
AD62	88	1
AD63	131	1

AD64	89	2
AD65	124	1
AD66	105	2
AD67	108	2
AD68	74	2

AUTHOR

Isabelle Poupard Santizo, Assistant Professor, Continuing Education, Defense Language Institute Foreign Language Center.

ALL INDEX

(2007-2017)

ARTICLES

- Ağazade, Ali Sidki. (2014). *Frequency versus Importance: Language Learning Strategy Use in the EFL Classroom*. 23-24(1&2), p. 47.
- Aiello, Jacqueline. (2014). *Variation in Second Language Learners' Strategies among Non-native English Speakers from Three Language/Culture Backgrounds*. 23-24(1&2), p. 23.
- Ayan, Hatice Ç. (2015). *(De)Motivation in Preparatory EFL Classrooms*. 25(1&2), p. 16.
- Bae, Gyseon. (2008). *What Causes Reliance on English? Challenges and Instructional Suggestions in a Drive for Using a Target Language Only*. 18(1&2), p. 63.
- Baker, Wendy. (2010). *Exploring Diverse Settings for Language Acquisition and Use: Comparing Study Abroad, Service Learning Abroad, and Foreign Language Housing*. 20(1&2), p. 45.
- Bohn, Mariko T. (2015). *Pedagogical Perspectives on Gendered Speech Styles in the Teaching and Learning of Japanese as a Foreign Language*. 25(1&2), p. 41.
- Bown, Jennifer. (2010). *Exploring Diverse Settings for Language Acquisition and Use: Comparing Study Abroad, Service Learning Abroad, and Foreign Language Housing*. 20(1&2), p. 45.
- Burnham, Kevin R. (2014). *Phonetic Training in the Foreign Language Curriculum*. 23-24(1&2), p. 63
- Byon, Andrew Sangpil. (2007). *Process-Driven Culture Learning in American KFL Classrooms*. 17(1&2), p. 73.
- Choi, Yujeong. (2014). *Hypothesis Testing in Task-Based Interaction*. 23-24(1&2), p. 1.
- Cohen, Andrew D. (2007). *Being a Strategic Language Learner in CALL*. 17(1&2), p. 57.
- Daloglu, Aysegul. (2007). *Teacher-Student Partnership in Evaluating and Revising a Multidisciplinary Sustained-Content English Language Course*. 17(1&2), p. 15.
- Danan, Martine. (2016). *Enhancing Listening with Captions and Transcripts: Exploring Learner Differences*. 26(2), p. 1.
- David, Virginia. (2016). *A Comparison of Two Approaches for Assessing L2 Writing: Process-based and Impromptu Timed Writing Exams*. 26(1), p. 65.

- DeSantis, Paulina. (2008). *Text Enhancement and the Acquisition of English Verbal Inflection -s by L1 Haitian Creole Speakers*. 18(1&2), p. 27.
- Dewey, Dan. (2010). *Exploring Diverse Settings for Language Acquisition and Use: Comparing Study Abroad, Service Learning Abroad, and Foreign Language Housing*. 20(1&2), p. 45.
- Diaz, Erin McNulty. (2017). *The Order of Explicit Information in Processing Instruction*. 27(1&2), p. 41.
- Eisenstein Ebsworth, Miriam. (2014). *Variation in Second Language Learners' Strategies among Non-native English Speakers from Three Language/Culture Backgrounds*. 23-24(1&2), p. 23.
- Felix-Brasdefer, J. Cesar. (2009). *Dispreferred Responses in Interlanguage Pragmatic Refusal Sequences in Learner - NS Interactions*. 19(1&2), p. 1.
- Gan, Zhengdong. (2015). *What Happens to Students' English after One Year of English-medium Course Study at University?* 25(1&2), p. 71.
- González-Bueno, Manuela. (2011). *The Teaching of Second Language Pronunciation through Processing Instruction*. 21(1 & 2), p.57.
- Graham, John. (2008). *Quantifying the Efficiency of a Translator: The Effect of Syntactical and Literal Written Translations on Language Comprehension Using the Machine Translation System FALCon (Foreign Area Language Converter)*. 18(1&2), p. 17.
- Graham, C. Ray. (2011). *The Effect of Explicit Instruction versus Exposure Only on ESL Learners' Ability to Recognize, Use, and Recall Phrasal Verbs*. 21(1&2), p. 31.
- Han, Hyunjoo. (2007). *Issues of Validity of SAT Subject Test: Korean with Listening*. 17(1&2), p. 33.
- He, Lei. (2011). *Metacognition in English as a Foreign Language Pronunciation Learning among Chinese Tertiary Learners*. 21(1&2), p. 1.
- Hedgcock, John. (2016). *Differentiating Heritage and Foreign Language Learners of Spanish: Needs, Perceptions, and Expectations*. 26(1), p. 1.
- Hernández, Todd A. (2008). *Integrative Motivation as a Predictor of Achievement in the Foreign Language Classroom*. 18(1&2), p. 1.
- Hernández, Todd A. (2016). *Short-Term Study Abroad: Perspectives on Speaking Gains and Language Contact*. 26(1), p. 39.
- Holman, Don. (2017). *Across the Threshold: A Call for ILR Proficiency Descriptor Bank*. 27(1&2), p. 1.
- Houston, Tony. (2006) *Communication Strategies in the Foreign Language Classroom*. 16(2), p. 65.
- Hubert, Michael. D. (2017). *Using Writing to Teach Pronunciation: An Experimental Fourth-Year University Spanish Phonetics/Phonology Course*. 27(1&2), p. 18.

- Jantzi, Jennifer K. (2008). *Quantifying the Efficiency of a Translator: The Effect of Syntactical and Literal Written Translations on Language Comprehension Using the Machine Translation System FALCon (Foreign Area Language Converter)*. 18(1&2), p. 17.
- Johnson, Cary. (2010). *Exploring Diverse Settings for Language Acquisition and Use: Comparing Study Abroad, Service Learning Abroad, and Foreign Language Housing*. 20(1&2), p. 45.
- Kennedy, Lauren. (2010). *The Reading Proficiency Interview (RPI): A Rapid Response Test Development Model for Assessing Reading Proficiency on the ILR Scale*. 20(1&2), p. 1.
- Kilpatrick, Cynthia. (2014). *Hypothesis Testing in Task-Based Interaction*. 23-24(1&2), p. 1.
- Kim, Eunju. (2008). *What Causes Reliance on English? Challenges and Instructional Suggestions in a Drive for Using a Target Language Only*. 18(1&2), p. 63.
- Koyama, Nobuko. (2016). *Introduction to a Dubbing Activity for a College First-Year Japanese Language Course*. 26(2), p. 69.
- Lee, Saekyun H. (2007). *Issues of Validity of SAT Subject Test: Korean with Listening*. 17(1&2), p. 33.
- Lefkowitz, Natalie. (2016). *Differentiating Heritage and Foreign Language Learners of Spanish: Needs, Perceptions, and Expectations*. 26(1), p. 1.
- Li, Yang. (2016). *The Effects of Pragmatic Consciousness-Raising Activities on the Learning of Speech Acts in the Beginning CFL Classroom*. 26(2), p. 53.
- Li, Shaofeng. (2009). *The Differential Effects of Implicit and Explicit Feedback on Second Language (L2) Learners at Different Proficiency Levels*. 19 (1&2), p. 53.
- Lim, Hye-Yeon. (2009). *Culture, Attributions, and Language Anxiety*. 19(1&2), p. 29.
- Lord, Gillian. (2007). *The Role of the Lexicon in Learning Second Language Stress Patterns*. 17(1&2), p. 1.
- Marsh, Fulya. (2007). *Teacher-Student Partnership in Evaluating and Revising a Multidisciplinary Sustained-Content English Language Course*. 17(1&2), p. 15.
- Martinsen, Rob A. (2010). *Exploring Diverse Settings for Language Acquisition and Use: Comparing Study Abroad, Service Learning Abroad, and Foreign Language Housing*. 20(1 & 2), p. 45.
- McCulloh, Ian A. (2008). *Quantifying the Efficiency of a Translator: The Effect of Syntactical and Literal Written Translations on Language Comprehension Using the Machine Translation System FALCon (Foreign Area Language Converter)*. 18(1 & 2), p. 17.
- Miller, Zachary F. (2016). *Military and Civilian L2 Instructors: Decoding Perceptions of U.S. Service Academy Cadets*. 16(2), p. 25.

- Morton, Jillian. (2008). *Quantifying the Efficiency of a Translator: The Effect of Syntactical and Literal Written Translations on Language Comprehension Using the Machine Translation System FALCon (Foreign Area Language Converter)*. 18(1&2), p. 17.
- Musayeva Vefali, Gülşen. (2014). *Frequency versus Importance: Language Learning Strategy Use in the EFL Classroom*. 23-24(1&2), p. 47.
- Musayeva Vefali, Gülşen. (2015). *(De)Motivation in Preparatory EFL Classrooms*. 25(1&2), p. 16.
- Oxford, Rebecca L. (2015). *How Language Learners Can Improve Their Emotional Functioning: Important Psychological and Psychospiritual Theories*. 25(1&2), p. 1.
- Podoprigora, Serguei A. (2010). *Passage Rating: L4 versus L5 Characteristics in the Russian Text Typology*. 20(1&2), p. 17.
- Quintana-Lara, Marcela. (2011). *The Teaching of Second Language Pronunciation through Processing Instruction*. 21(1&2), p. 57.
- Razavi, Nikta. (2014). *Variation in Second Language Learners' Strategies among Non-native English Speakers from Three Language/Culture Backgrounds*. 23-24(1&2), p. 23.
- Rodriguez, Amy M. (2008). *Quantifying the Efficiency of a Translator: The Effect of Syntactical and Literal Written Translations on Language Comprehension Using the Machine Translation System FALCon (Foreign Area Language Converter)*. 18(1&2), p. 17.
- Santizo, Isabelle Poupard. (2017). *Correlating Aptitude with Oral Proficiency: A Quantitative Study of DLAB and OPI Scores Across Four Language Categories*. 27(1&2), p. 73.
- Stansfield, Charles, W. (2010). *The Reading Proficiency Interview (RPI): A Rapid Response Test Development Model for Assessing Reading Proficiency on the ILR Scale*. 20(1&2), p. 1.
- Stapleton, Paul. (2015). *What Happens to Students' English after One Year of English-medium Course Study at University?* 25(1&2), p. 71.
- Tang, Frank Lixing. (2014). *Variation in Second Language Learners' Strategies among Non-native English Speakers from Three Language/Culture Backgrounds*. 23-24(1&2), p. 23.
- Vigil, Donny. *Using Writing to Teach Pronunciation: An Experimental Fourth-Year University Spanish Phonetics/Phonology Course*. 27(1&2), p. 18.
- Yang, Chi Cheung. (2015). *What Happens to Students' English after One Year of English-medium Course Study at University?* 25(1&2), p. 71.
- Yeon, Sang-Hee. (2008). *Training English Word-Final Palatals to Korean Speakers of English*. 18(1&2), p. 51.
- Yeon, Sang-Hee. (2014). *It Works for Me: From the High Achievers' Perspective*. 23-24(1&2), p. 75.

Zhu, Jia. (2016). *The Effects of Pragmatic Consciousness-Raising Activities on the Learning of Speech Acts in the Beginning CFL Classroom*. 26(2), p. 53.

REVIEWS

Forsyth, Edo. (2010). *Review: Political Russian: An Intermediate Course in Russian Language for International Relations, National Security and Socio-Economics*. 20(1&2), p. 74.

Hedgcock, John. (2008). *Review: Conversation Analysis and Language for Specific Purpose*. 18(1&2), p. 77.

Henery, Ashlie. (2011). *Review: Language Learning and Study Abroad: A Critical Reading of Research*. 21(1&2), p. 93.

Hoffman, Mika. (2010). *Review: Text Complexity and Reading Comprehension Tests*. 20 (1&2), p. 67.

Hur, HyunSoo. (2015). *Review: Language Teacher Education for a Global Society*. 25(1&2), p. 92.

Hur, HyunSoo. (2016). *Review: Discourse Analysis: Putting Our Worlds into Words*. 26(2), p. 94.

Jackson, Gordon L. (2011). *Review: Methodology and Materials Design in Language Teaching: Current Perceptions and Practices and their Implications*. 21(1&2), p. 83.

Kelley, Jeremy. (2007). *Review: Leon and Bhatt: Structure du Francais Moderne: Introduction a l'analyse linguistique (3rd Ed. Rev.)*. 17(1&2), p. 91.

Lee, Mina. (2014). *Review: Creativity and Innovation in Language Education*. 23-24(1&2), p. 91.

Martel, Jason. (2014). *Review: Reflective Practice in ESL Teacher Development Groups: From Practices to Principles*. 23-24(1&2), p. 87.

Martel, Jason. (2016). *Review: Adult Learning in the Language Classroom*. 26(2), p. 89.

Nassif, Lama. (2010). *Review: Spoken Corpora in Applied Linguistics*. 20(1&2), p. 70.

Thomas, Murray R. (2016). *Review: Postsecondary Foreign Language Teaching in the United States*. 26(1), p. 83.

UPCOMING EVENTS 2017

JANUARY

- January 5-8* Linguistic Society of American (LSA) Annual Meeting, Austin, TX.
Information: www.linguisticsociety.org
- January 5-8* Modern Language Association (MLA) Convention, Philadelphia, PA.
Information: www.mla.org/convention

FEBRUARY

- February 2-5* American Association of Teachers of Slavic and East European Languages (AATSEEL), San Francisco, CA.
Information: www.aatseel.org
- February 16-19* California Language Teachers' Association Annual Conference, Monterey, CA.
Information: <http://clta.net/future-conferences/>
- February 26-27* 19th International Conference on Linguistics, Language Teaching and Learning, Barcelona, Spain.
Information:
www.waset.org/conference/2017/02/barcelona/ICLLTL/home

MARCH

- March 9-11* Central States Conference on the Teaching of Foreign Languages (CSCTFL), Chicago, IL.
Information: www.csctfl.org
- March 16-18* Southern Conference on Language Teaching (SCOLT), Orlando, FL.
Information: www.scolt.org
- March 18-21* American Association for Applied Linguistics (AAAL), Portland, OR.
Information: www.aal.org
- March 21-24* Teachers of English to Speakers of Other Languages (TESOL) International Convention, Seattle, WA.
Information: www.tesol.org

APRIL

April 27 – May 1 American Educational Research Association (AERA) Annual Meeting, San Antonio, TX.
Information: www.era.net

MAY

May 16-18 Computer-Assisted Language Instruction Consortium (CALICO) annual conference. Flagstaff, AZ.
Information: calico.org

May 28-June 2 NAFSA: Association of International Educators Annual Conference and Expo, Los Angeles, CA.
Information: www.nafsa.org

JUNE

June 15-17 International Society for Language Studies (ISLS) Annual Conference, Honolulu, HI.
Information: www.isls.co/index.html

JULY

July 16-19 American Association of Teachers of French (AATF) 2016 conference, St Louis, MO.
Information: www.frenchteachers.org

NOVEMBER

November 18-21 Middle East Studies Association (MESA) Annual Meeting, Washington, DC.

Information: mesana.org/annual-meeting/upcoming.html

November 17-19 American Council on the Teaching of Foreign Languages Annual Convention (ACTFL), Nashville, TN.

Information: www.actfl.org

November 17-19 Chinese Language Teachers Association (CLTA) Annual Conference, Nashville, TN.

Information: clta-us.org

November 17-19 American Association of Teachers of German (AATG) Annual Conference, Nashville, TN.

Information: www.aatg.org

November 17-19 American Association of Teachers of Japanese (AATJ) Fall Conference, Nashville, TN.

Information: www.aatj.org

INFORMATION FOR CONTRIBUTORS

Submission Information for Authors

AIMS AND SCOPE

Applied Language Learning (ALL) is to promote professional communication within the Defense Language Program and academic communities on adult language learning for functional purposes.

The Editor encourages the submission of research and review manuscripts from such disciplines as: (1) instructional methods and techniques; (2) curriculum and materials development; (3) testing and evaluation; (4) implications and applications of research from related fields in linguistics, education, communication, psychology, and social sciences; and (5) assessment of needs within the profession.

SPECIFICATIONS FOR MANUSCRIPTS

Prepare the manuscripts in accordance with the following requirements:

- Follow the American Psychological Association (APA) style (the 6th Edition)
- Not exceeding 6,000 words (not including reference, appendix, etc.)
- Use double spacing, with margins of one inch on all four sides
- Use Times New Roman font size 12
- Number all pages consecutively
- In black and white only, including graphics and tables
- Create graphics and tables in a Microsoft Office application (such as Word, PowerPoint, Excel)
- Graphics and tables should not exceed 4.5” in width
- Do not use the footnotes and endnotes function in MS Word. Insert a number formatted in superscript following a punctuation mark. Type notes on a separate page
- Keep the layout of the text as simple as possible

SUBMISSION REQUIREMENT

Applied Language Learning publishes only original works that have not been previously published elsewhere and that are not under consideration by other publications.

Each submission must contain (1) a title page, including author information; (2) abstract of the article; (3) five keywords; and (4) manuscript, including references.

Send all submissions electronically to the Editor: jiaying.howard@dliflc.edu.

REVIEW PROCESS

Manuscripts will be acknowledged by the editor upon receipt and subsequently sent out for peer review. Authors will be informed about the status of the article once the peer reviews have been received and processed. Reviewer comments will be shared with the authors. Once an article has been accepted for publication, the author will receive further instructions regarding the submission of the final copy.

CORRESPONDENCE

Send all inquiries and editorial correspondence by email to the Editor:

jiaying.howard@dliflc.edu.

Guidelines for Manuscript Preparation

RESEARCH ARTICLE

Divide your manuscript into the following sections, in the order listed below:

1. Title and Author Information
2. Abstract
3. Keywords
4. Text body, including:
 - Acknowledgements (optional)
 - Notes (optional)
 - References
 - Tables and figures (optional)
 - Appendixes (optional)

REVIEW ARTICLE

It should describe, discuss, and evaluate publications that fall into a topical category in foreign language education. The relative significance of the publications in the context of teaching realms should be pointed out. A review article should be 15 to 20 double-spaced pages.

REVIEW

Submit reviews of textbooks, scholarly works on foreign language education, dictionaries, tests, computer software, audio-video materials, computer and mobile applications, and other non-print materials. Point out both positive and negative aspects of the work(s) being considered. In the three to five double-spaced pages of the manuscript, give a clear but brief statement of the work's content and a critical assessment of its contribution to the profession. Keep quotations short. Do not send reviews that are merely descriptive.

COMMENTARY

ALL invites essays that exchange ideas and views on innovative foreign language education, and comments on matters of general academic or critical interest or on articles in previous issues. Essays should not exceed 2,000 words.

CALL FOR PAPERS

Applied Language Learning, a refereed journal published semiannually by the Defense Language Institute Foreign Language Center and Presidio of Monterey, is soliciting articles for publication.

The Journal (*US* ISSN 1041-679X and ISSN 2164-0912 for the online version) is to provide a forum for the exchange of ideas and information on instructional methods and techniques, curriculum and materials development, assessment of needs within the profession, testing and evaluation, and implications and applications of research from related fields such as linguistics, education, communications, psychology, and the social sciences. The journal seeks to serve the professional interest of language teachers, administrators, and researchers concerned with the teaching of foreign languages to adult learners. We welcome articles that describe innovative and successful practice and methods and/or report educational research or experimentation.

Deadline: Submissions are welcome at any point. Manuscripts received by 31 March will be considered for the spring issue and by 30 September for the fall issue of the journal.

Send your manuscript electronically to the Editor:

jjaying.howard@dliflc.edu

Read the recent and past issues of *Applied Language Learning* at:
<http://www.dliflc.edu/resources/publications/applied-language-learning/>