

Educational Linguistics

Paula Winke
Susan M. Gass *Editors*

Foreign Language Proficiency in Higher Education

[Click here for link to full book!](#)

 Springer

Editors

Paula Winke
Second Language Studies Program
Michigan State University
East Lansing, MI, USA

Susan M. Gass
Second Language Studies Program
Michigan State University
East Lansing, MI, USA

ISSN 1572-0292

ISSN 2215-1656 (electronic)

Educational Linguistics

ISBN 978-3-030-01005-8

ISBN 978-3-030-01006-5 (eBook)

<https://doi.org/10.1007/978-3-030-01006-5>

Library of Congress Control Number: 2018961737

© Springer Nature Switzerland AG 2019

This work is subject to copyright. All rights are reserved by the Publisher, whether the whole or part of the material is concerned, specifically the rights of translation, reprinting, reuse of illustrations, recitation, broadcasting, reproduction on microfilms or in any other physical way, and transmission or information storage and retrieval, electronic adaptation, computer software, or by similar or dissimilar methodology now known or hereafter developed.

The use of general descriptive names, registered names, trademarks, service marks, etc. in this publication does not imply, even in the absence of a specific statement, that such names are exempt from the relevant protective laws and regulations and therefore free for general use.

The publisher, the authors, and the editors are safe to assume that the advice and information in this book are believed to be true and accurate at the date of publication. Neither the publisher nor the authors or the editors give a warranty, express or implied, with respect to the material contained herein or for any errors or omissions that may have been made. The publisher remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

This Springer imprint is published by the registered company Springer Nature Switzerland AG
The registered company address is: Gewerbestrasse 11, 6330 Cham, Switzerland

A Cross-Linguistic and Cross-Skill Perspective on L2 Development in Study Abroad



Dan E. Davidson and Jane Robin Shaw

Abstract The present study reports on measured gains in L2 proficiencies in speaking, reading and listening of U.S. students ($N = 308$) who took part in year-long federally funded overseas immersion programs for Arabic, Chinese and Russian. Subjects were late adolescent and young adult learners of diverse social and economic backgrounds participating in year-long structured instructed immersion programs hosted in China, Kazakhstan, Moldova, Morocco and Russia. L2 gains in post-program proficiency levels from 4.76 to 7.74 standard deviations above pre-program measured levels are reported for both the early- and the late-stage learners: Mean post-program proficiency levels of ILR-2, CEFR-B2 are demonstrated by the early-stage learners across skills in all three target languages. The mean post-program proficiency levels of ILR-3, CEFR-C1 of the university subjects meets certification levels for language-designated positions in in most U.S. government and professional organizations. The study also examines skill gains across modalities: Advanced participants show concurrent gains across three skills: reading, listening, and speaking. Post-program reading and speaking are strongly correlated with pre-program listening at the advanced levels. Reading ability is strongly associated with gains in speaking and in listening skills, as the student progresses from novice through the professional level.

Keywords L2 gain · Immersion · Study abroad · Cross-skill correlations · Critical languages · SLA · Diversity abroad · Professional proficiency

D. E. Davidson (✉)

American Councils Research Center (ARC), American Councils for International Education, Washington, DC, USA

Myra T. Cooley Lectureship, Bryn Mawr College, Bryn Mawr, PA, USA

e-mail: davidson@americancouncils.org

J. R. Shaw

Department of Russian, Bryn Mawr College, Bryn Mawr, PA, USA

e-mail: jshaw@brynmawr.edu

© Springer Nature Switzerland AG 2019

P. Winke, S. M. Gass (eds.), *Foreign Language Proficiency in Higher Education*, Educational Linguistics 37, https://doi.org/10.1007/978-3-030-01006-5_12

217

At the request of the U.S. Congress, the American Academy of Arts and Sciences (AAAS) has released a major report on language learning in the United States, *America's Languages: Investing in Language Education for the 21st Century* (AAAS, 2017). Among five areas recommended for policy attention in the Academy Report was a call for expanded access to study abroad, for “students to travel, experience other cultures, and immerse themselves in languages as they are used in everyday interactions and across all segments of society” (AAAS, 2017, p. 27).

As noted in the Academy Report, language learning in the study abroad context has the capacity to produce significant linguistic and cultural gains, but overseas study is also costly, and substantial growth in language is by no means achieved by all those who go abroad (Freed, 1998; Mason, Powers, & Donnelly, 2015; Vande Berg, Connor-Linton, & Paige, 2009). As with any educational setting, program design, teacher preparation, student motivation, time-on-task, and an appropriately supportive environment for learning are critical components for successful language acquisition in the study abroad context. While the total number of U.S. students who study abroad has increased over the past two decades to 313,415 annually, most study currently takes place in English-speaking regions (*Open Doors*, 2016). Moreover, despite the well-documented benefits of longer-term immersion, only 2.5% of Americans studying abroad in 2014–2015 stayed a full academic year, reflecting an unfortunate decline in long-term study over the past twenty years (Dwyer, 2004; Kinginger, 2011; Pellegrino Aveni, 2005). In a recent large-scale comparison of summer and academic-year overseas study programs, language gains were compared for differing target languages, initial levels, and program durations: The greatest gains, regardless of starting point or target language, were associated with year-long programs (Davidson, 2015).

For those students who do undertake serious year-long language study, the structured, federally-sponsored programs initiated under the National Security Language Initiative of 2006 (see, for example, <https://exchanges.state.gov/us/program/nsliy>) have demonstrated a capacity over the past decade for producing advanced and superior-level speakers on the American Council on the Teaching of Foreign Languages (ACTFL) proficiency scale (see ACTFL, 2012). Alumni of the NSLI-Y (National Security Language Initiative Youth) high school senior secondary program (SP) are fully prepared to enter advanced-level university course work taught in the target language, and alumni of the undergraduate Flagship programs (UP; see <https://thelanguageflagship.org/>) regularly go on to join the U.S. workforce as bilingual professionals (Murphy & Evans-Romaine, 2015; Powell & Lowenkron, 2006; USED, 2008). Unfortunately, the number of Americans benefiting from these programs still falls far short of meeting the growing needs of government, business, international development, and society at large (Brecht, Rivers, Robinson, & Davidson, 2015; Damari, Rivers, Brecht, Gardner, & Robinson, 2017; Rivers, 2012).

1 Assessing Language Acquisition in the Study Abroad Context

The systematic study of language acquisition during study abroad is now a well-recognized subfield of second language acquisition scholarship (cf. special issue of *System*, 2017; Winke & Gass, 2018). Within that subfield, issues of variation in learning outcomes and ultimate attainment dominate much of the empirical research (Baker-Smemoe, Dewey, Bown, & Martinsen, 2014; Brecht, Davidson, & Ginsberg, 1995; Davidson, 2015; DeKeyser, 2007; Dewey, 2004; Freed, 1998; Mason et al., 2015; Watson, Siska, & Wolfel, 2013). Group-level analyses of standard measures of pre-program and post-program L2 proficiencies using the ILR (<http://www.gov-tilr.org/>) or ACTFL scales for speaking, reading, and listening can provide a validated and reliable cross-linguistic basis for comparing relatively robust numbers of student records, while controlling statistically for target language, modality, initial L2 level, duration of immersion, and other linguistic and learner background variables.

Proficiency-based measures are widely used today as a component of programs of formative and summative assessment as well as for participant certification purposes. The ACTFL and ILR proficiency scales, and the larger *World-readiness standards for learning languages* (NSFLEP, 2015) of which they are a part, are widely used today in K-12 (e.g., the Seal of Biliteracy in 25 states), college placement, study abroad, and teacher education programs (ACTFL, 2012; NSFLEP, 2015). Most importantly, the ILR scale, on which the ACTFL proficiency guidelines are based, is used virtually across all U.S. government agencies recruiting for language-designated positions (Herzog, n.d.; Interagency Language Roundtable, 2016; Liskin-Gasparro, 1984; Tschirner, 2011). As widely used as these standardized measures are, the authors acknowledged inherent limitations with these and other large-scale L2 proficiency models and testing scales currently in use, whether ILR/ACTFL, the Common European Framework of Reference, the TORFL (Russian), TOPIK (Korean), HSK (Mandarin), or others (see Bachman, 1988; Kramsch, 2014; North, 2006). While the current generation of proficiency tests do not capture the full dynamic range of linguistic and cultural repertoires of which the L2 user may be capable, the final proficiency rating, whether for speaking, writing, reading, or listening, is based on an individualized analysis of evidence produced by the candidate under controlled examination conditions: an L2 product (interpersonal communication, presentation, textual interpretation) evaluated in terms of its overall effectiveness and appropriateness for the intended audience.

2 Assessing L2 Across Modes and Modalities

Previous research on cross-skill gains in language proficiency in the study abroad environment has noted a relationship at the intermediate-to-advanced levels between study abroad participants' initial reading levels and their ultimate gains in listening comprehension, while strong structural control of the L2 has been consistently associated with gains across all skills (Brecht et al., 1995). Grammatical (structural) knowledge makes both visual and aural input comprehensible and allows the language learner to improve processing speed and build confidence in reading and listening (Krashen, 1985; Norris & Ortega, 2003; Ortega, 2009). Examining factors affecting L2 gain at the "superior" (ILR 3) level, Davidson (2010) observed that pre-program listening comprehension levels among advanced students of Russian were predictive of program-final oral proficiency outcomes. The higher the initial listening comprehension score, the more likely the candidate was to achieve a score of 3 ("superior") or higher in speaking by the end of the academic program. Noting the need for further study of cross-skill correlations with oral proficiency gains, the author hypothesized that strong listening comprehension appears to be critical for the L2 learner in detecting and acting on the spoken feedback of native interlocutors (e.g. re-tracings, recasts, paraphrasing) that make up a regular part of the students' extended interactions with local friends and contacts in the immersion context at that level (Davidson, 2010; Winke & Gass, 2018). Self-correction behavior, more generally, has also been identified as yet another predictor of L2 gain for young adults in the immersion environment (Golonka, 2000).

The American Academy report identifies several federally-sponsored programs as exemplifying best practices in effective overseas immersion language training (AAAS, 2017). Two of those federal programs, one open to senior secondary students (SP), the other primarily to undergraduate juniors and seniors (UP), monitor student progress through systematic pre- and post-program assessment of proficiency levels in speaking, reading, and listening. The two federal programs will serve as sources of performance-based data for the present comparative study of learning outcomes for American students of Arabic, Chinese, and Russian in the overseas immersion context. In addition to skill-specific reports for each of the target languages, cross-skill relations by skill-specific proficiency levels will be reported and compared here as well.

3 Study Participants

The present study includes data from late-adolescent and young adult participants in two major federal programs focused on an intensive in-country immersion study of Arabic, Chinese, and Russian ($N = 308$). They include year-long undergraduate students (UP) of Russian ($N = 126$) and pre-college participants in the Arabic academic-year program ($N = 47$), Chinese academic-year program ($N = 78$), and

Russian academic-year program ($N = 57$). The federal funding model for these programs was designed to encourage participation by students from a greater range of socioeconomic backgrounds than is normally possible for fee-based study abroad programs. The SP admits students on a competitive basis and without regard to their ability to pay, including students with no prior experience of learning the target language in question. The UP subsidizes a substantial portion of all program costs—under certain circumstances, all costs—and requires applicants to demonstrate advanced level (ILR-2) proficiency in speaking and at least one other skill and to test at no lower than a 1+ in the third skill. A writing proficiency test is now being added to the testing portfolio for the UP languages but is not included in the present analysis. (See Appendix 1 and 2 for full list of selection criteria for both programs.)

While it is impossible to control for pre-selection effects in the analysis of these two cohorts, the researchers believe, given the basis on which candidates were selected and funded, that the outcomes data included here may be regarded as generally representative of the impact of a year of overseas language immersion study on that segment of the U.S. student population who elect to apply for and accept positions in a federally-funded study abroad program, regardless of their socioeconomic and educational backgrounds.¹

4 Data and Testing Instruments

Testing, conducted at the beginning and end of the programs, includes face-to-face or telephonic oral proficiency interviews (OPIs) and online proficiency-based reading and listening comprehension examinations based on the ACTFL and ILR scales (ACTFL, 2012; Interagency Language Roundtable Scale, 2016).

Anonymized participant score reports that included OPI pre- and post-program test scores were made available to the researchers and analyzed for a total of 308 year-long participants in Arabic, Chinese, and Russian. Reading and listening pre- and post-program scores were made available to the researchers for all advanced-level (UP) study subjects and for the Russian study subset of the early-stage (SP) learners. The overseas study programs in question took place between 2009 and 2014 in China, Jordan, Kazakhstan, Moldova, Morocco, Russia, and Taiwan. Small-group instruction, peer tutors, homestays, attendance of regular local classes, and integrated cultural enhancement programs were standard features of all programs. UP students also participated in internships. Detailed description of the overseas program designs and interventions in use over the 2009–2014, including

¹Analysis of the distribution of K-12 foreign language enrollments across the U. S. indicates a correlation between socio-economic levels within a school district (as reflected in the 2010 U.S. Census) and the likelihood that the district will (or will not) offer a foreign language, defined as a language other than English, at the K-12 level. See discussion of estimate models, *National FL Enrollment Survey* (2017), <https://www.americancouncils.org/sites/default/files/FLE-report-June17.pdf>

the two selected for analysis in the present study (SP and UP), may be found in Davidson (2015).

The combined speaking, reading, and listening comprehension data (reflecting presentational, interpersonal, and interpretive modes of communication) are considered by the authors to provide a robust, cross-modal, multi-language array of aligned measures of L2 skills for use in assessing and comparing outcomes for the programs under study. Pre-program scores reflect the skill levels participants brought to their study-abroad experiences and serve as a baseline against which subsequent changes are measured.

5 Research Questions

Given the recognized value of overseas immersion for the acquisition of foreign languages at the advanced and professional levels, reliable information on learning outcomes across different target languages and with regard to specific skills should be widely available to teachers, advisors, and policymakers concerned with the preparation of a new generation of L2 users and professionals. For that reason, the present study poses the following research questions:

1. What are the mean gains in oral proficiency of students who participate in structured year-long study programs in Arabic, Chinese, and Russian, as measured by changes in levels of proficiency as well as in units of pre-program standard deviation?
2. To what extent do the choice of target language and the student's initial levels of proficiency affect gains in OPI?
3. To what extent are second language (L2) gains in reading, listening, and speaking correlated for students in the year-long study-abroad context? To what extent do specific pre-program skills account for post-program attainment across skills?

The growth in importance of study abroad and in access to study abroad has made these questions more broadly relevant today than was the case in years past. The current study hopes to advance understanding of linguistic factors that contribute to successful learning of three critical languages, Arabic, Chinese, and Russian, in the study abroad context.

6 Data Collection and Preparation

OPI testing was conducted by ACTFL-certified oral proficiency testers; post-program OPI testing of participants with higher initial proficiency levels was administered face-to-face, while pre-testing and lower-range tests were administered telephonically. Proficiency-based reading and listening comprehension tests were administered under proctored, computer-mediated conditions. Test specifications,

item development, and scoring protocols for the reading and listening proficiency tests are described in detail by Bazarova, Lekic, and Marshall (2009); statistical documentation of the reading and listening tests may be found at Wothke and Petersen (2017).

Testing data are reported using the ACTFL and/or the ILR scales, depending on the phase of study of the participant. For ease of statistical analysis, the researchers converted ILR-scaled scores to ACTFL scores using the following conversions:

0+ to novice-high,
1 to intermediate-mid,
1+ to intermediate-high,
2 to advanced,
2+ to advanced high,
3 to superior,
3+ to superior-high,
4 to distinguished.

To avoid introducing additional measurement error as a result of the necessary score conversion, and given that the ACTFL and ILR scales do not fully align, the authors also report ILR data in those cases where score conversions were undertaken. This procedure is consistent with other recent studies (Davidson, 2015; Davidson, Garas, & Lekic, 2016; Mason, Powers, & Donnelly, 2015).

Since proficiency scores represent ordinal values, pre- and post-program score columns in the data sets with numeric values were then created. Integers from 1 to 18 were assigned for each ACTFL rating from novice-low to distinguished, with novice-low as 1. An additional unit was added to the coding to account for threshold-level junctures on the proficiency scale (novice, intermediate, advanced, superior, and distinguished). Thus, novice-high to intermediate-low is marked by a move from 3 to 5, while intermediate-high to advanced-low is represented by a numerical shift from 7 to 9 on the linear scale, and so forth. While none of the participants received a final program score of “superior-low,” as testers do not generally give this score, a space of 1 unit was left in the column for this rating in order to maintain consistent intervals across languages and proficiency levels. Values for all ACTFL ratings and for the intervening values are given below in Table 1:

Given the nature of the three-dimensional construct, the “inverted pyramid,” employed for ILR and ACTFL proficiency assessment, L2 gains tend to post at a more rapid rate at lower levels of proficiency but require increasingly more time as the participant grows and advances to higher levels of proficiency (Brecht et al., 1995). A more nuanced mathematical model has yet to be developed and accepted within the foreign language assessment community to account statistically for the time-on-task differentials implicit in the successive levels of the ILR proficiency scale. (See also Tigchelaar, this volume, for additional information on this.) The use (above) of an additional numerical value (4, 8, 12, 16) at each threshold level along the scale is an entirely arbitrary but statistically helpful intervention both to mark the additional functional and expressive capacity represented by the next level up on the scale and to mitigate the effects of restriction of range within clusters of pre- and post- test scores.

Table 1 Numerical values by ACTFL (Ordinal) ratings

ACTFL rating	Value in database
Novice-low	1
Novice-mid	2
Novice-high	3
<i>(Threshold)</i>	4
Intermediate-low	5
Intermediate-mid	6
Intermediate-high	7
<i>(Threshold)</i>	8
Advanced-low	9
Advanced-mid	10
Advanced-high	11
<i>(Threshold)</i>	12
<i>(Superior-low)</i>	13
Superior-mid	14
Superior-high	15
<i>(Threshold)</i>	16
<i>(Distinguished-low)</i>	17
Distinguished	18

7 Analysis

To assess language specific and overall L2 gain within the immersion programs, initial distributions were run of OPI values for each language (Arabic, Chinese, and Russian) using data from all year-long SP and UP participants ($N = 459$). The distributions were categorized by pre-program OPI and post-program OPI; score changes (“delta” values), if any, were tabulated and included for each as well. The subjects were divided into three groups for analytic purposes: those who began the program at the “novice” proficiency level, those who began at the “intermediate” level, and those who began at the “advanced” level.

To test for relationships across modalities, multivariate pairwise correlations were run and univariate simple statistics were recorded using data from SP and UP Russian participants. (Reading and listening data were available only for the Russian subset of SP but for all participants in UP.) Both Pearson and Spearman correlations were generated. For each participant grouping (“novice,” “intermediate,” and “advanced”), a set of correlations among delta (reading), delta (listening), and delta (OPI) was generated (with “delta” signifying change within scores from pre-test to post-test); a set of correlations across all participant levels was also run.

Multivariate pairwise correlations, with corresponding univariate simple statistics, were also generated to test for relationships among initial and post-program levels in reading and listening. Sets of correlations were run across all levels and for each participant grouping (“novice,” “intermediate,” and “advanced”); pre-program

reading values, pre-program listening values, post-program reading values, and post-program listening values were correlated.

Both cross-skill and same-skill correlations (e.g. pre-/post-reading; pre-/post-listening, pre-/post-speaking) were performed throughout to verify the overall homogeneity of the data and to check, in particular, for any significant differential effects that might influence the analysis related to participant gender, age, heritage background, program year, and program site. No significant external or programmatic effects were found (Shaw, 2017).

A third set of multivariate pairwise correlations and univariate simple statistics was generated to test for relationships among initial skills in reading and listening and post-program OPI attainment. Sets of correlations were run for each language and across all proficiency levels (“novice,” “intermediate,” and “advanced”); pre-program reading values, pre-program listening values, and post-program OPI values were tested.

In order to clarify further the relationship of language gains across skills (speaking, reading, and listening), distributions of gains with participants categorized, as previously, by pre-program OPI levels were run across modalities. Mean delta (skill) values for “novice,” “intermediate,” and “advanced” Russian academic-year SP and UP participants were considered and compared. While previous distributions focusing on OPI results included participants without reading and listening data, for this test, only participants with delta values in every modality and all three levels, the Russian-only data set, were considered.

Pre- and post-program reading and listening data and their respective relationships to OPI gains were run for each modality (reading and listening), adjusted for pre-program values for that modality. For one set of distributions, participants were grouped by novice, intermediate, and advanced pre-program reading values; mean pre-program reading values, post-program reading values, delta (reading), post-program OPI values, and delta (OPI) were considered. For a second set of distributions, participants were grouped by novice, intermediate, and advanced pre-program listening values; mean pre-program listening values, post-program listening values, delta (listening), post-program OPI values, and delta (OPI) were considered.

Regression analyses using fit models were performed on all year-long participant data to test statistical relationships between pre-program skills and ultimate OPI attainment as measured by post-program OPI values. For fit model type, standard least squares with emphasis on effect leverage were chosen. The results were represented as leverage plots, and corresponding statistics were generated.

Post-program OPI values represented the dependent variable. Plots with pre-program reading, pre-program listening, and pre-program OPI values as independent variables were generated. The plots and accompanying statistics were then examined to determine which independent variable had least effect and whether any variables had negative effects; new sets of leverage plots were then generated as applicable using the remaining variables.

8 Results

The present study has addressed the measurement of L2 gain across languages with respect to the student's initial level of proficiency and choice of target language; delta (OPI) and delta (OPI) in units of pre-program standard deviation were also calculated. The subject population, as noted above, comprised late-adolescent and young-adult learners of the critical languages. For all distributions, the duration of the immersion program (intervention) was one academic year (9 calendar months).

The effect of the immersion intervention on the cohort ($N = 77$) beginning the programs in Arabic, Chinese, and Russian at the novice level is highly significant, ranging from 6.36 (Arabic) to 6.91 (Chinese) to 7.30 (Russian) standard deviations above the measured pre-program means. For those beginning the program in the same three languages at the intermediate level ($N = 53$), the effect is again highly significant, but slightly weaker: Arabic (6.93), Chinese (4.76), and Russian (5.74). For those beginning the program at the advanced level ($N = 112$), the mean gain deltas are 7.74 for all participants.

Reviewing the three language-specific cohorts across programs, one notes that the proficiency gains (deltas) are comparable across all proficiency levels (as are the standard deviations), with gains at the intermediate level slightly more modest than those posted by the novices and advanced students. The latter is particularly significant in light of the expected effect of the measurement artifact, noted above.

9 Pre-/Post-program L2 Gain Levels and Gain Amounts (Deltas) by Modality

Multivariate analyses were conducted for participants across proficiency levels and at each specific level comparing delta (skill) values across modalities. For Russian academic-year participants (the only group for which speaking, reading, and listening proficiency scores were available for all levels of study), Pearson correlations showed a moderate, statistically significant correlation between gains in reading and in listening over the period of study, noted here as "delta (R)" and "delta (L)" ($r = 0.3338$, $p = 0.0010$). Spearman correlations showed a moderate, statistically significant positive correlation between delta (R) and delta (L) ($\rho = 0.4002$, $p < 0.0001$).

For novice Russian academic-year participants, Spearman correlations showed a strong, statistically significant positive correlation between delta (R) and delta (L) values ($\rho = 0.6825$, $p = 0.0207$). For intermediate Russian academic-year participants, no correlations met the probability threshold for statistical significance. For advanced Russian academic-year participants, Pearson and Spearman correlations showed moderate, statistically significant positive correlations for all pairings. Spearman correlations also showed statistically significant positive correlations for all pairings, slightly stronger than Pearson but still moderate.

Academic year L2 Russian participant data ($N = 183$), from the cohort for which pre-and post-program reading and listening data were available for entering novice-level participants, as well as data for those who entered study at the intermediate and advanced proficiency levels) demonstrated strongly correlated, statistically significant relationships across all proficiency levels.

9.1 *Russian Academic-Year Participants, All Levels: Reading to Listening*

			Pearson correlations		Spearman correlations	
			Correlation	Sign. Prob.	Spearman rho	Prob. > rhol
Pre-R	–	Post-L	0.8882	<.0001*	0.6023	<.0001*
Pre-L	–	Post-R	0.8914	<.0001*	0.7344	<.0001*

Pre-program reading levels were strongly correlated with post-program listening outcomes, and, conversely, pre-program listening levels also predicted post-program reading attainment.

9.2 *Intermediate Participants: Reading to Listening*

			Pearson correlations		Spearman correlations	
			Correlation	Sign. Prob.	Spearman rho	Prob. > rhol
Pre-R	–	Post-L	0.9412	<.0001*	0.9396	<.0001*
Pre-L	–	Post-R	0.8969	<.0001*	0.8930	<.0001*

The finding is consistent with Brecht et al. (1995), which focused exclusively on semester-length overseas Russian immersion, and Davidson (2010), which compared summer, semester, and academic year outcomes for overseas Russian. In both studies, reading proficiency was strongly correlated, in turn, with target-language grammatical/structural control, and both reading and grammar served as more or less equivalent predictors of ultimate oral proficiency gain at the advanced level.

9.3 Pre-reading/Pre-listening to Post-OPI

Multivariate analyses for pre-reading (pre-R) and pre-listening (pre-L) scores with post—program OPI results for *all levels* of study showed the following results:

			Pearson correlations		Spearman correlations	
			Correlation	Sign. Prob.	Spearman rho	Prob. > rhol
Pre-R	–	Post-OPI	0.7896	<.0001*	0.5956	<.0001*
Pre-L	–	Post-OPI	0.8041	<.0001*	0.6700	<.0001*

Pearson correlations were slightly stronger for pre-program listening and post-program OPI ($r = 0.8041$, $p < 0.0001$) than pre-program reading and post-program OPI ($r = 0.7896$, $p < 0.0001$) for the cohort as a whole.

10 Comparison of Mean AY Skill Gains (R, L, S) by Initial OPI Proficiency Level

Based on existing program data, for which academic-year immersion data are available at all three levels, the distribution of skill-specific gains categorized by the participant's initial (pre-program) speaking proficiency presents the following results.

	Delta (Reading)		Delta (Listening)		Delta (OPI)	
	Mean	Std. Dev.	Mean	Std. Dev.	Mean	Std. Dev.
Novice	3.20	2.10	3.50	1.65	5.70	1.70
Intermediate	4.23	1.79	4.00	1.52	4.54	1.27
Advanced	3.72	1.76	3.86	1.50	3.99	1.13

Novice (N = 10), Intermediate (N = 13), Advanced (N = 69)

As seen in the earlier distribution of Russian academic-year participants by level (Table 3.2), the mean gain delta for the cohort of advanced-level students is slightly smaller numerically than for those who began the program with speaking levels at the novice or intermediate level, most likely an effect of the measurement artifact discussed above. Looking at all three levels, however, it is clear that *the immersion experience for early stage learners is accompanied by relatively rapid rates of gain in speaking*. As the learners' speaking skills improve, the data show a more evenly distributed range of skill gains (both means and the size of standard deviations). Increased ability and opportunities for self-expression and interactions with locals also multiply the need for cross-skill and multi-modal forms of communication at the intermediate and advanced levels.

The results of analyses of pre-and post-program reading levels and their respective relationships to OPI gains are presented in the following distributions:

10.1 Distributions of all AY Scores Based on Initial Levels of Reading Comprehension

		Novice pre-program reading	Intermediate pre-program reading	Advanced pre-program reading
Pre-program reading	Mean	1.80	5.44	10.41
	Std. Dev.	0.42	0.73	0.50
Post-program reading	Mean	6.00	9.00	14.57
	Std. Dev.	2.49	1.87	1.27
Delta (Reading)	Mean	4.20	3.56	4.16
	Std. Dev.	2.35	1.67	1.39
Post-program OPI	Mean	8.50	9.44	14.13
	Std. Dev.	1.35	1.51	1.36
Delta (OPI)	Mean	5.00	4.78	4.16
	Std. Dev.	1.49	1.72	1.27

Novice pre-program reading (N = 10), Intermediate pre-program reading (N = 9), Advanced pre-program reading (N = 63)

Novice-level readers showed the greatest gains in both reading and in speaking, followed by those who began the program as advanced-level readers. Gains in reading were notable for each group, ranging from 3.56 to 4.20 mean delta (reading). When delta (OPI) values were compared based on participants' pre-program reading levels, mean delta (OPI) decreased only very slightly from 5.00 (novice) to 4.78 (intermediate) to 4.16 (advanced), differences most likely resulting from the effects of the measurement artifact itself.

The results of analyses of pre-and post-program listening levels and their respective relationships to OPI gains are presented in the following distributions:

10.2 Distributions of all AY Scores Based on Initial Levels of Listening Comprehension

		Novice pre-program listening	Intermediate pre-program listening	Advanced pre-program listening
Pre-program listening	Mean	2.46	6.50	10.46
	Std. Dev.	0.52	0.71	0.50
Post-program listening	Mean	6.15	11.60	14.37
	Std. Dev.	1.68	2.72	0.89
Delta (Listening)	Mean	3.69	5.10	3.91
	Std. Dev.	1.65	2.13	0.95
Post-program OPI	Mean	8.46	11.30	14.22
	Std. Dev.	1.45	2.00	1.22
Delta (OPI)	Mean	4.92	4.50	4.17
	Std. Dev.	1.75	0.97	1.21

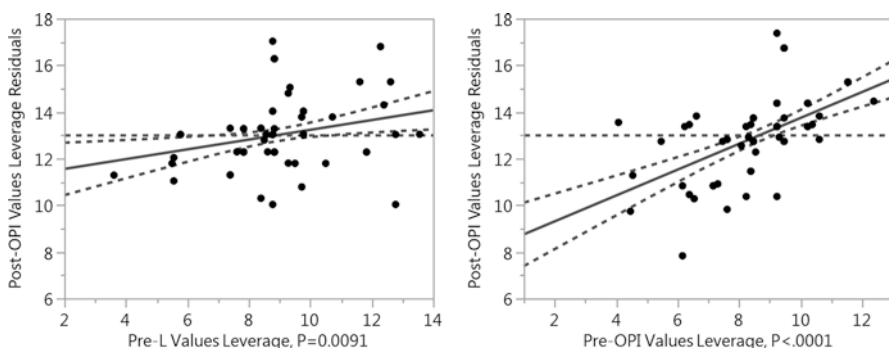
Novice pre-program listening (N = 13), Intermediate pre-program listening (N = 10), Advanced pre-program listening (N = 65)

Unlike for speaking and reading, participants who began the immersion program with intermediate levels of pre-program listening showed the greatest gains over the course of the immersion year. That said, listening gains were substantial for all groups, ranging from 3.69 to 5.10 mean delta (listening). When delta (OPI) values were compared based on participants' pre-program listening levels, mean delta (OPI) were observed to decrease very slightly from novice to intermediate to advanced listeners. This mirrors the patterns for delta (OPI) gains by level previously noted for speaking and reading. In fact, no matter which pre-program skill was selected as the independent variable, very similar patterns in delta (OPI) appeared, trends which are programmatically significant precisely because they are so small in this case, given the well-documented effects of the measurement artifact itself.

11 Fit Model Analysis: Pre-program Levels as Predictors of Post-program OPI

When the effects of the three pre-program variables were tested, parameter estimates were 0.02162 for pre-program reading, 0.1924 for pre-program listening, and 0.5487 for pre-program OPI. Pre-program reading had a slight positive effect but did not meet the threshold for statistical significance. Pre-program listening had a positive effect on post-program OPI; while it did not meet the threshold for statistical significance, it came much closer than did pre-program reading. Pre-program OPI had a substantial, statistically significant positive effect on post-program OPI.

However, when pre-program reading, the independent variable with the least leverage on post-program OPI, was removed and the relative effects of pre-program listening and pre-program OPI and analyzed again, the results were notable:



Parameter estimates were 0.2094 for pre-program listening and 0.5544 for pre-program OPI; both independent variables had positive, statistically significant effects on post-program OPI. Pre-program listening was observed to contribute more than 20% of the variation in post-program OPI values present in the model.

12 Findings and Discussion

Overall, the immersion intervention effects for the early-stage SP learners and for the advanced-level SP and UP subjects were highly significant: 7.30 and 7.74 standard deviations respectively. The linguistic and cultural impact on both groups is significant, permitting graduates of SP to enter college-level courses at sophomore and junior levels and UP graduates to move directly into government and private sector positions requiring professional levels (ILR-3) of linguistic and intercultural competence and above.

Mean delta (OPI) was 5.81 for novice participants, 4.36 for intermediate participants, and 4.19 for advanced participants; gains in units of pre-program standard deviation were 7.30 for novice participants, 5.58 for intermediate participants, and 7.74 for advanced participants. Proficiency gains were comparable across levels, with gains at the intermediate level slightly more modest than those posted by the novices and advanced students. The latter is noteworthy in that advanced-level gains are relatively more difficult to achieve, due to the effects of the measurement artifact (the inverted pyramid), which assumes considerably greater effort and time-on-task to move from Level 2 to Level 3 than from 1 to 2 or from 0 to 1 (ACTFL, 2012; Brecht et al., 1995). As noted above, the expected decline in mean delta (OPI) values as proficiency levels rise was, in fact, very gradual and barely observable. These findings relate to presumed cognitive, academic, and socio-emotional effects of the immersion intervention at more advanced levels of acquisition on the learning process.

A set of distributions was generated later in the study comparing delta (OPI), delta (reading), and delta (listening) values among Russian academic-year participants; these distributions included only those participants for whom delta (skill) values in all three modalities were available. Consideration of the mean delta (OPI) data for this participant subset allows the examination of patterns in gains by level with slightly different selection criteria in place. As noted above, given the well-documented “artifact effect” of the proficiency measurement model (the inverted pyramid), mean deltas would be expected to decrease as the student progresses in learning along the proficiency scale from one proficiency threshold to the next highest. For these cohorts, however, the delta values were still relatively robust for every level and modality: 5.70 for novice participants, 4.54 for intermediate participants, and 3.99 for advanced participants. As with the previous set of distributions for Russian academic-year participants, the change in mean delta (OPI) with increasing level was indeed observable, but limited. Cross-testing of other participant groups within the larger database did not substantially alter this pattern. The consistency of these results may be seen to further attest to the value of overseas language immersion as a facilitator of language gain at all levels, and to its particular value at the upper-intermediate and advanced levels, when comparable gains are more difficult to achieve in the domestic learning context.

13 Cross-Modality Patterns for Study-Abroad Participants

To assess the relationships among pre-program skill levels and post-program outcomes within Russian academic-year program, several sets of multivariate pairwise correlations were run. When gains across all levels were examined, Pearson correlations showed a moderate, statistically significant correlation between delta (reading) and delta (listening) ($r = 0.3338$, $p = 0.0010$). Spearman correlations showed a moderate, statistically highly significant positive correlation between delta (reading) and delta (listening) ($\rho = 0.4002$, $p < 0.0001$) and a moderately weak but

statistically significant positive correlation between delta (reading) and delta (OPI) ($\rho = 0.2361$, $p = 0.0220$). Reading and listening gains increased together across levels, as did reading and OPI gains.

When limiting the examination of skill gains to novice-level students within the Russian academic-year programs, reading and listening gains were found to increase together at the novice level. For advanced academic-year participants, Pearson correlations showed moderate, statistically significant positive correlations for delta (reading) and delta (OPI), $r = 0.2423$ and $p = 0.0449$; and for delta (listening) and delta (OPI), $r = 0.2592$ and $p = 0.0315$. Whether examined via parametric or non-parametric correlations, delta (listening) and delta (OPI) showed a slightly stronger relationship than delta (reading) and delta (OPI) among advanced participants. Earlier research has noted a relationship between listening and OPI among students of Russian at the advanced level and above (Davidson, 2010).

As a further exploration of the relationships among gains in different modalities, a set of distributions was generated comparing delta (OPI), delta (reading), and delta (listening) values among academic-year participants. In a comparison of different levels relative to one another, for delta (reading) and delta (listening), data for the intermediate-level cohort showed the greatest gains.

When the delta values for different skills of participants at a given level were considered, certain trends appeared: for the novices, delta (OPI) was a great deal higher than delta (reading) and delta (listening). For the intermediates and advanced, delta (OPI) was only slightly higher. For the advanced group, all the delta values were relatively similar; this close correspondence among delta values seems to mirror the consistent pattern of positive delta (skill) correlations seen among advanced participants in the overseas immersion program setting, a tendency towards the equalization of skill differentials in the context of the full immersion, acquisition-rich environment.

In addition to the examination of relationships among delta (skill) values, pre-program and post-program values across modalities were also analyzed. Possible cross-modal patterns in reading and listening were investigated via correlations of pre-program reading, pre-program listening, post-program reading, and post-program listening values among Russian academic-year participants. Of particular note, pre-program reading had a notably strong positive relationship with post-program listening ($r = 0.9412$ and $p < 0.0001$ with Pearson correlations, $\rho = 0.9396$ and $p < 0.0001$ with Spearman correlations). Similarly, pre-program listening had a strong positive relationship with post-program reading ($r = 0.8969$ and $p < 0.0001$ with Pearson correlations, $\rho = 0.8930$ and $p < 0.0001$ with Spearman correlations). The relationship between pre-program reading and post-program listening was slightly stronger than the relationship between pre-program listening and post-program reading.

For novice Russian academic-year participants, pre-program reading to post-program listening and pre-program listening to post-program reading were both positively correlated; pre-program reading to post-program listening was slightly more strongly correlated. For intermediate Russian academic-year participants, all categories were positively correlated. As with novices, pre-program reading to

post-program listening was slightly stronger than pre-program listening to post-program reading.

For advanced participants, in contrast, pre-program listening to post-program reading was positively correlated, while pre-program reading to post-program listening was not. As has been previously noted, “novice,” “intermediate,” and “advanced” participant categories have been delineated by pre-program OPI for testing purposes. While OPI levels serve as a good measurement of participants’ overall L2 proficiency level, certain participants enter programs with relatively greater differences in a skill other than speaking, and, thus, may be seen to straddle category borders from a cross-modal testing perspective. Upper-level academic-year participants, who normally represent a greater period of previous study of the target language, have presumably experienced a broader range of instructional styles and a more diverse array of language-learning approaches by skill. As observed above in the analysis of delta (skill) values, advanced academic-year program participants show relatively similar and proportionate degrees of gain across all modalities while enrolled overseas.

A final series of analyses was conducted to examine the relationship of pre-program reading and pre-program listening values to post-program attainment as represented by post-program OPI values. For academic-year participants across all levels, both pre-program reading and pre-program listening were strongly correlated with post-program OPI whether examined via parametric or non-parametric correlations. For pre-program reading and post-program OPI, $r = 0.7896$ and $p < 0.0001$ with Pearson correlations, while for pre-program listening and post-program OPI, $r = 0.8041$ and $p < 0.0001$. Of the two pre-program skills in question, pre-program listening showed a modestly stronger correlation to post-program OPI.

For both novices and intermediates, pre-program listening and post-program OPI were highly correlated. Pre-program reading and post-program OPI were also correlated but fell short of the threshold for significance for either participant group.

From the point of view of the foreign language teacher or supervisor, the practical conclusion that flows from the relationship between pre-program listening and post-program OPI may be to recognize the importance of developing listening comprehension at the earliest stages of study. The observation of a correlation between listening comprehension and OPI gain at the intermediate level has not previously been reported in the literature.

14 Distributions as a Measurement of L2 Gain Across Modalities in Russian

As part of a consideration of gains in modalities beyond OPI, a set of distributions of Russian academic-year participants grouped by their pre-program reading levels was run. These included pre-program reading, post-program reading, and delta (reading) values for each skill-specific participant level as well as delta (OPI) and

post-program OPI. A second set of distributions was run with participants grouped by their pre-program listening levels; contents included pre-program listening, post-program listening, and delta (listening) values for each level as well as delta (OPI) and post-program OPI.

Fit group model analyses were conducted to examine the relationship of language gains across modalities. Pre-program values in all skills (reading, listening, and OPI) were leveraged to see how much they each accounted for gains as represented by post-program OPI results. When all three pre-program variables were examined jointly, parameter estimates were 0.02162 for pre-program reading ($p = 0.8376$), 0.1926 for pre-program listening ($p = 0.0970$), and 0.5487 for pre-program OPI ($p < 0.0001$). As expected, effects within the same modality were pronounced: pre-program OPI had the greatest effect on post-program OPI results, representing more than 50% of the variable portion explained by the model. Of the two cross-modal categories, pre-program listening approached the threshold for statistical significance and contributed a notable amount of the variable portion of the model. In contrast, pre-program reading did not contribute meaningfully to the overall effect. To further explore the strength of the effect of pre-program listening on post-program OPI results and tighten the model, pre-program reading was removed.

When the test was rerun with pre-program listening and pre-program OPI as the two independent variables, parameter estimates were 0.2094 for pre-program listening ($p = 0.0091$) and 0.5544 for pre-program OPI ($p < 0.0001$). Pre-program listening accounted for 21% of the variance in post-program OPI results, thus demonstrating a strong cross-modality effect.

15 Conclusions

The present study reports on L2 outcomes (measured changes in L2 proficiency levels in speaking, reading, and listening) of U.S. students ($N = 308$) who took part in year-long federally funded overseas immersion programs for Arabic, Chinese, and Russian. The subjects of the study were late adolescent and young adult learners, selected through a competitive process for participation in a group of well-resourced and carefully monitored year-long structured immersion programs at established host-country institutions in China, Kazakhstan, Moldova, Morocco, Russia, and Taiwan. The target languages in question represent a group of languages deemed “critical” for U.S. national security and economic interests by the U.S. government and considered typologically “difficult” (linguistically and in terms of time-on-task learning requirements) for English base-language learners (Thompson, 2014) in comparison to more commonly taught foreign languages, such as French, German, or Spanish.

The authors make no claim regarding the generalizability of these findings for study abroad programs, other than for those year-long models which have provided data for the present study. However, the notably high levels of language gain (rang-

ing from 4.1 to 7.1 standard deviations above the measured pre-program proficiency levels) reported here for both the early- and the late-stage students of critical languages have both policy and practical implications for the modern language profession and for all those concerned with preparing a new generation of graduates for a workforce in which professional-level language and intercultural skills are increasingly in demand (Brecht et al., 2015; Rivers, 2015).

The mean post-program proficiency levels (ACTFL/Advanced, CEFR-B2) demonstrated by the early-stage learners (SP) across skills are sufficient to ensure those students successful placement into advanced-level target-language courses offered at most U.S. universities (American Councils for International Education, 2017a; Bärenfänger & Tschirner, 2012). The mean post-program proficiency levels (ACTFL/Superior, CEFR-C1, C2) of the UP graduates represented in the study correspond to the professional language competencies required of those seeking employment in language-designated positions in many government agencies, as well as for those who expect to make use of their language skills in academia, business, research, international development, or domestic social services. Participants in both the early-stage (SP) and the advanced-level (UP) cohorts registered similar threshold-level L2 gains, regardless of the choice of critical language. In this context, it should be noted that while UP participants were required to meet an ILR-2 (ACTFL/Advanced) qualifying level in at least two skills at the time of application to the program, while early-stage learners were accepted at both the intermediate and novice levels of proficiency. Indeed, approximately one third of the entering students in SP reported no knowledge of the L2 prior to participation in the overseas programs.

Participant language gains are well-correlated across modalities. Advanced participants show concurrent gains across three skills: reading, listening, and speaking. Post-program reading and listening are strongly correlated, in turn, with pre-program listening skills. Initial levels of listening comprehension (pre-listening score) are positively correlated with growth in speaking skills at the intermediate and advanced levels, while reading ability, which functions as a proxy measure for more general levels of L2 structural and lexical control, is strongly associated with gains in speaking and in listening abilities, as the student progresses from novice to intermediate and to the advanced levels.

Of further note in the present study is empirical evidence of a process of *cross-skill equalization* as learners progress to the advanced and superior levels, despite notable early-stage skill gaps at the novice and intermediate levels among these groups. (Heritage learners are not included in the present study.)

Established practice within the foreign language field has focused on the value of study abroad for American L2 students who have completed one to three years of prior formal study, either in school or at the university level. The practice is understandable if study abroad is viewed as a one-time, relatively expensive intervention in (or enhancement of) the student's domestic undergraduate learning career. However, the latest survey/census of K-16 foreign language enrollments in the United States unfortunately confirms that no more than 20% of pupils currently have access to foreign language classes in U.S. school districts, while fewer than 7% of those who attend college enroll in a foreign language course (Brecht et al., 2013;

Brecht et al., 2015; American Councils for International Education, 2017b; Goldberg, Looney, & Lusin, 2015). Hence, requirements for prior study of the language as a prerequisite for study abroad exclude far too large a segment of the U.S. population to meet minimal standards of fairness and equal opportunity, even when issues of cost are put aside. The present study provides evidence of the notable language learning success that U.S. students of all backgrounds and with little or no prior study of an L2 can achieve in the overseas structured immersion context. Within the course of one year, students acquire levels of functional proficiency that can be put to immediate use in academia, service sectors, internships, and in their future careers.

Study abroad is a recognized “high-impact” practice in U.S. higher education (Kuh, 2012, 2016), and language-empowered study abroad can produce substantial linguistic gains for late-adolescent and young-adult learners across modalities, as demonstrated here, gains not typical in most domestic settings (Carroll, 1967; Kuh, 2012, 2016; Tschirner, 2011). In light of the declining rates of U.S. undergraduate participation in longer-term, language-focused study abroad (noted above), the present study offers further empirical support for the recent call by the American Academy of Arts and Sciences to foreign language departments, study abroad advisors, and institutional leaders to expand opportunities for language study at all levels in the context of institutionally approved education abroad activities, supported as well by the major federal initiatives aimed at preparing a new generation of linguistically and culturally competent U.S. professionals.

Appendices

Appendix 1 (SP)

NSLI for Youth Eligibility Requirements

www.nsliforyouth.org

NSLI-Y programs offer intensive language immersion in a variety of locations around the world. Scholarships are available for students to learn the following languages: Arabic, Bahasa Indonesia Chinese (Mandarin), Hindi, Korean, Persian (Tajiki), Russian, and Turkish.

Programs may take place in the following locations: China, Estonia, India, Indonesia, Jordan, Korea, Latvia, Moldova, Morocco, Russia, Taiwan, Tajikistan, Turkey and other locations around the world.

Eligibility Requirements

- U.S. citizen
- Grade point average (GPA) of 2.5 or higher on a 4.0 scale, or the equivalent

- 15–18 years of age at start of program (birthdate between July 10, 1999 and June 10, 2003 for summer programs; birthdate between September 20, 1999 and June 30, 2003 for academic year programs)
- Enrolled in high school (including home school)
- Not an immediate family member of an employee of the U.S. Department of State who works in the Youth Programs Division of the Bureau of Educational and Cultural Affairs or an employee at a NSLI-Y administering organization whose duties involve the NSLI-Y program
- Have not previously traveled outside the U.S. on a long-term (more than eight weeks) program sponsored by the Bureau of Educational and Cultural Affairs, Department of State
- Previous NSLI-Y *summer* program participants or participants of ECA-funded short-term programs are only eligible to apply for a NSLI-Y academic year program.

Previous language study is not a requirement. Students of all levels of language ability are encouraged to apply.

The NSLI-Y program seeks applicants who represent the diversity of the United States. Students of all racial, ethnic, religious, gender identities, sexual orientations, and socio-economic backgrounds are welcome to apply, as are students with disabilities.

Appendix 2 (UP)

The Language Flagship Capstone Program

www.thelanguageflagship.org

The Flagship Capstone full-year immersion is open to all Domestic Flagship undergraduate students who are committed to attaining professional or superior-level language proficiency through an intensive language training program tailored to their professional interests and academic specialization. It may occur during the third, fourth, or fifth year of a student's undergraduate program. The model also assumes and encourages that, in addition to full-year study, students will complete an additional period of immersion overseas to accelerate their language learning.

Applicants should have a strong academic record, a demonstrated interest in advancing their Arabic, Russian, Persian, Chinese, Hindi/Urdu, Korean, Portuguese, and Turkish skills and using these languages in their future career, and a desire to share their understanding of this language and culture within the larger community.

Undergraduate Applicants

All students who are enrolled at one of the Domestic Flagship Programs and reach the required proficiency level ILR-2 in their language on an Oral Proficiency Interview (OPI) and at least on one of the online modalities (reading, listening, writing), while scoring no lower than level ILR-1+ on the remaining two online modalities, are accepted to the Overseas Program, upon recommendation of the Overseas Project Directors.

Russian Overseas Flagship Post-BA or “At-Large” Applicants

The Russian Overseas Flagship Program accepts qualified applicants who did not participate in a Domestic Flagship Program and already have a bachelor’s degree. The participants are selected on the basis of their language skills, academic merits, previous experience of study abroad, and ability to demonstrate how advanced Russian skills are going to help their career plans. At-large applicants to the Russian Overseas Flagship Program must either possess a B.A. degree or expect to receive one before starting the program. Successful applicants who are not heritage speakers must have completed at least three years of language at a college level and must have participated in a language study program in a Russian-speaking country for at least six weeks.

References

- ACTFL. (2012). *ACTFL proficiency guidelines 2012*. Alexandria, VA. Retrieved from https://www.actfl.org/sites/default/files/pdfs/public/ACTFLProficiencyGuidelines2012_FINAL.pdf
- American Academy of Arts and Sciences. (2017). *America’s languages: Investing in language education for the 21st century*. Cambridge, MA: Author. Retrieved from <https://www.amacad.org/content/Research/researchproject.aspx?i=21896>
- American Councils for International Education. (2017a). *National examination in world languages (NEWL)*[®]. Proficiency score levels and placement recommendations: <https://www.americancouncils.org/services/testing-and-assessment/NEWL>
- American Councils for International Education. (2017b). *National K-16 foreign language enrollment survey*. N. Garas, Project Director. Retrieved from <https://www.americancouncils.org/sites/default/files/FLE-report-June17.pdf>
- Bachman, L. (1988). Problems in examining the validity of the ACTFL oral proficiency interview. *Studies in Second Language Acquisition*, 10(2), 149–164. <https://doi.org/10.1017/S0272263100007282>
- Baker-Smemoe, D., Dewey, D. P., Bown, J., & Martinsen, R. A. (2014). Variables affecting L2 gains during study abroad. *Foreign Language Annals*, 47(3), 464–486. <https://doi.org/10.1111/flan.12093>
- Bärenfänger, O., & Tschirner, E. (2012). *Assessing Evidence of Validity of Assigning CEFR Ratings to the ACTFL Oral Proficiency Interview (OPI) and the Oral Proficiency Interview by computer (OPIc) (Technical Report 2012-US-PUB-1)*. Leipzig, Germany: Institute for Test Research and Test Development.

- Bazarova, S., Lekic, M. D., & Marshall, C. (2009). The online proficiency-based reading, listening, and integrated writing external assessment program for Russian: A report to the field. *Russian Language Journal*, 59, 59–78.
- Brecht, R., Davidson, D., & Ginsberg, R. (1995). Predictors of foreign language gain during study abroad. In B. F. Freed (Ed.), *Second language acquisition in a study abroad context* (pp. 37–66). Amsterdam, Netherlands: John Benjamins.
- Brecht, R., Abbott, M., Davidson, D., Rivers, W. P., Robinson, J., Slater, R., & Yoganathan, A. (2013). Languages for all? In *The Anglophone challenge*. College Park, MD: University of Maryland.
- Brecht, R. D., Rivers, W. P., Robinson, J. R., & Davidson, D. E. (2015). Professional language skills: Unprecedented demand and supply. In T. Brown & J. Bown (Eds.), *To advanced language proficiency and beyond: Theory and methods for developing superior second-language ability* (pp. 171–184). Washington, DC: Georgetown University Press.
- Carroll, J. B. (1967). Foreign language proficiency levels attained by language majors near graduation from college. *Foreign Language Annals*, 1, 131–151. <https://doi.org/10.1111/j.1944-9720.1967.tb00127.x>
- Damari, R. R., Rivers, W. P., Brecht, R. D., Gardner, C. P., & Robinson, J. (2017, February 28). The demand for multilingual human capital in the U. S. labor market. *Foreign Language Annals*, 50, 13.
- Davidson, D. E. (2010). Study abroad: When, how long, and with what results? New data from the Russian front. *Foreign Language Annals*, 43(2), 6–26.
- Davidson, D. E. (2015). The development of L2 proficiency and literacy within the context of the federally supported overseas language training programs for Americans. In T. Brown & J. Bown (Eds.), *To advanced language proficiency and beyond: Theory and methods for developing superior second-language ability* (pp. 117–150). Washington, DC: Georgetown University Press.
- Davidson, D. E., Garas, N., & Lekic, M. D. (2016). Assessing language proficiency and intercultural development in the overseas immersion context. In D. Murphy & K. Evans-Romaine (Eds.), *Exploring the US language flagship program: Professional competence in a second language by graduation* (pp. 156–176). Bristol, UK: Multilingual Matters.
- DeKeyser, R. M. (2007). Study abroad as foreign language practice. In R. M. DeKeyser (Ed.), *Practice in a second language: Perspectives from applied linguistics and cognitive psychology* (pp. 208–226). Cambridge, UK: Cambridge University Press.
- Dewey, D. P. (2004). A comparison of reading development by learners of Japanese in intensive domestic immersion and study abroad contexts. *Studies in Second Language Acquisition*, 26(2), 303–327. <https://doi.org/10.1017/S0272263104262076>
- Dwyer, M. M. (2004). More is better: The impact of study abroad program duration. *Frontiers: The Interdisciplinary Journal of Study Abroad*, 10, 151–163.
- Freed, B. F. (1998). An overview of issues and research in language learning in a study abroad setting. *Frontiers: The Interdisciplinary Journal of Study Abroad*, 4(2), 31–60.
- Goldberg, D., Looney, D., & Lusin, N. (2015). *Enrollments in languages other than English at US institutions of higher education Fall 2013*. Modern Language Association. https://www.mla.org/content/download/31180/1452509/EMB_enrllmnts_nonEngl_2013.pdf
- Golonka, E. M. (2000). *Identification of salient linguistic and metalinguistic variables in the prediction of oral proficiency gain at the advanced-level threshold among adult learners of Russian*. Unpublished dissertation, Bryn Mawr College.
- Herzog, M. (n.d.). *An overview of the history of the ILR language proficiency skill level descriptions and scale*. Retrieved from <http://govtilr.org/Skills/IRL%20Scale%20History.htm>
- Interagency Language Roundtable. (2016). *ILR skill level descriptions*. Retrieved from <http://govtilr.org/>
- Kinginger, C. (2011). Enhancing language learning in study Abroad. *Annual Review of Applied Linguistics*, 31, 58–73.

- Kramsch, C. (2014). Teaching foreign languages in an era of globalization: Introduction. *The Modern Language Journal*, 98(1), 296–311. <https://doi.org/10.1111/j.1540-4781.2014.12057.x>
- Krashen, S. D. (1985). *The input hypothesis: Issues and implications*. London, UK: Longman.
- Kuh, G. D. (2012). High-impact educational practices: What they are, who has access to them, and why they matter. *Peer Review*, 14(3), 29.
- Kuh, G. D. (2016). *Study Abroad as a high-impact practice: Retrospective and prospective*. Retrieved from http://www.ifs-a-butler.org/images/stories/pdf/advisors/Kuh_IFSA-Butler_6-18-2016.pdf
- Liskin-Gasparro, J. E. (1984). The ACTFL proficiency guidelines: A historical perspective. In T. V. Higgs (Ed.), *Teaching for proficiency, the organizing principle* (pp. 11–42). Lincolnwood, IL: National Textbook.
- Mason, L., Powers, C., & Donnelly, S. (2015). *The Boren awards: A report of oral language proficiency gains during academic study abroad*. New York, NY: Institute of International Education.
- Murphy, D., & Evans-Romaine, K. (Eds.). (2015). *Exploring the US language flagship program. Professional competence in a second language by graduation*. Bristol, UK: Multilingual Matters.
- Norris, J., & Ortega, L. (2003). Defining and measuring SLA. In C. J. Doughty & M. H. Long (Eds.), *The handbook of second language acquisition* (pp. 717–761). Malden, MA: Blackwell. <https://doi.org/10.1002/9780470756492.ch21>
- North, B. (2006). *The common European framework of reference: Development, theoretical and practical issues*. Paper presented at the Japan-Europe International Symposium: A New Direction in Foreign Language Education: The Potential of the Common European Framework of Reference for Languages, Osaka University of Foreign Studies, Japan.
- NSFLEP. (2015). *World-readiness standards for learning languages* (4th ed.). Alexandria, VA: The American Council on the Teaching of Foreign Languages (ACTFL).
- Open Doors*. (2016). *Report on international educational exchange online*. Retrieved from <http://www.iie.org/Research-and-Publications/Open-Doors/Data>
- Ortega, L. (2009). *Understanding second language acquisition*. London, UK: Hodder Education.
- Pellegrino Aveni, V. A. (2005). *Study abroad and second language use: Constructing the self*. Cambridge, UK: Cambridge University Press.
- Powell, D., & Lowenkron, B. (2006). *National security language initiative*. Washington, DC: Office of the Spokesman. Retrieved from <https://2001-2009.state.gov/r/pa/prs/ps/2006/58733.htm>
- Rivers, W. P. (2012). The unchanging American capacity in languages other than English: Speaking and learning languages other than English, 2000–2008. *Modern Language Journal*, 96(3), 369–379.
- Rivers, W. P. (2015). *The contributions of language to the economic interests of the United States*. Prepared by the Joint National Committee for Languages (JNCL). <https://www.amacad.org/multimedia/pdfs/TheContributionsOfLanguagetoeconomicInterestsOftheUnitedStates.pdf>
- Shaw, J. R. (2017). *Understanding language gain in the overseas immersion context: Multi-modal assessment of young adult learners of Arabic, Chinese, and Russian* (Doctoral dissertation). Bryn Mawr College, Bryn Mawr, PA.
- Thompson, I. (2014). *Language learning difficulty*. Retrieved from <http://aboutworldlanguages.com/language-difficulty>
- Tschirmer, E. (2011). Reasonable expectations: Frameworks of reference, proficiency levels, educational standards. *Studies in Applied Linguistics* (1), 101–119. (Revised English version of Vernünftige Erwartungen: Referenzrahmen, Kompetenzniveaus, Bildungsstandards, 2008, *Zeitschrift für Fremdsprachenforschung*, 19(2), 187–208).
- USED. (2008). *Enhancing foreign language proficiency in the United States: Preliminary results of the National Security Language Initiative*. Washington, DC: U.S. Department of Education. Available at <https://nsep.gov/sites/default/files/nsli-preliminary-results.pdf>

- Vande Berg, M., Connor-Linton, U. J., & Paige, M. R. (2009). The Georgetown Consortium project: Interventions for student learning abroad. *Frontiers: The Interdisciplinary Journal of Study Abroad*, 18, 1–75.
- Watson, J. R., Siska, P., & Wolfel, R. L. (2013). Assessing gains in language proficiency, cross-cultural competence, and regional awareness during study abroad: A preliminary study. *Foreign Language Annals*, 46(1), 62–79.
- Winke, P., & Gass, S. (2018). When some study abroad: How returning students align with the curriculum and impact learning. In C. Sanz & A. Morales-Front (Eds.), *Handbook of study Abroad research and practice* (pp. 527–543). New York, NY: Routledge.
- Wothke, W. & Petersen, K. (2017). *Technical specifications. American Councils Language Assessment Systems (ACLASS)*. Available at <https://www.americancouncils.org/technical-reports>

Dan E. Davidson is Director of American Councils Research Center (ARC) in Washington, D. C. and Emeritus Professor of Russian and Second Language Acquisition at Bryn Mawr College, USA. He has written extensively in the fields of language, culture and educational development, including a 20-year longitudinal study of L2 acquisition during study abroad. His recent study, “Assessing Language Proficiency and Intercultural Development in the Overseas Immersion Context,” appeared in *Exploring the US Language Flagship. Professional L2 Competence by Graduation* (Multilingual Matters).

Jane Robin Shaw is Lecturer within the Department of Russian and the Russian Language Institute of Bryn Mawr College, USA, where she defended her Ph.D. dissertation, “Understanding Language Gain in the Overseas Immersion Context: Multi-Modal Assessment of Young Adult Learners of Arabic, Chinese, and Russian,” in 2017. Dr. Shaw’s scholarly interests include late adolescent and young adult SLA in the immersion context, the articulation of domestic and overseas L2 study, the concept of “language difficulty,” and the L2 assessment.