26

The Role of Cognitive Aptitudes in a Study Abroad Language-Learning Environment

Medha Tare, Ewa Golonka, Alia K. Lancaster, Carrie Bonilla, Catherine J. Doughty, R. Kirk Belnap, and Scott R. Jackson

Introduction

Study abroad is an important training context for language learning and has been considered crucial for reaching advanced levels of language knowledge (Freed, 1995). Studies have shown superior performance by students who studied abroad compared to those who only studied at home across different skills and abilities (Llanes, 2011). Greater increases for participants who studied abroad have been shown in several studies for foreign language oral fluency (Freed, 1995; Segalowitz & Freed, 2004) as well as vocabulary (Dewey, 2008; Foster, 2009). Many factors have been implicated in students' language-learning success during study abroad, with proficiency gains found to be heavily dependent on time in country (Davidson, 2010; Larson-Hall & Dewey, 2012) and other noncognitive predictors, including gender, age, starting proficiency, and knowledge of other foreign languages (Brecht, Davidson, & Ginsberg, 1993; Davidson, 2010). Aptitudes, defined as individual differences in learners' stable cognitive abilities, may also relate to the effectiveness of a study abroad experience; however, the research on the role of aptitudes (Llanes, 2011) is inconclusive, and what has been conducted is limited by the sample size and breadth of aptitude constructs that were investigated. In this study, we examine the relationships between oral proficiency gains during study abroad and a wide range of aptitude constructs, as measured by the High-Level Language Aptitude Battery (Hi-LAB) of 10 cognitive measures.

Previous Literature

One area of interest in the study abroad literature is the role of individual differences in cognitive aptitude in determining students' success in language learning. This literature suggests that there is a relationship between learners' cognitive aptitudes and linguistic gains in the study abroad context, though the results thus far are mixed. A group of predictor studies used aptitude, as measured by the Modern Language Aptitude Test (MLAT; Carroll & Sapon, 1959), in an effort to identify factors affecting linguistic gains resulting from study abroad. For instance, using data from many years of study abroad programs in Russia, Brecht et al. (1993) examined predictors of foreign language gain during study abroad (including gender, prior language experience, knowledge of grammar, and skill proficiency) and scores on three sections of the MLAT. Regression analyses of 658 students' data showed that MLAT Sections 3 (English vocabulary and phonetic coding) and 4 (grammatical sensitivity) did predict gains in reading and listening scores; however, MLAT tests did not predict students' Oral Proficiency Interview (OPI) scores.

More recently, Larson-Hall and Dewey (2012) studied 44 learners' language proficiency in Japanese during mission trips abroad. The aptitude tests included subtests from the MLAT and LLAMA (Meara, 2005), and a nonword repetition task as a measure of phonological short-term memory; the outcomes were a Simulated OPI and an elicited imitation task in which learners repeated Japanese sentences. In a regression model that included the aptitude tests, motivation, and the time spent on the mission, results showed that for both outcome variables, time spent on the mission accounted for the greatest variance, but the phonological short-term memory aptitude measure still accounted for 14%–15% of the variance in learners' outcomes. Although MLAT and LLAMA measures were predictive in a basic analysis, only the nonword repetition measure was also predictive in the complete analysis. Thus, while this study and prior research showed a limited relationship between MLAT and study abroad outcomes, other cognitive measures of aptitude, such as working memory, have proven more likely to predict success in study abroad/immersion contexts.

A recent meta-analysis showed that working memory plays a supportive role in both second-language processing and development of proficiency (Linck, Osthus, Koeth, & Bunting, 2014). Several studies have examined more specifically how individual differences in this ability interact with the demands that the study abroad context places on processing real-time language input. For instance, O'Brien, Segalowitz, Freed, and Collentine (2007) investigated the role of phonological memory in Second Language (L2) oral fluency development in study abroad and at-home contexts. Students, who had at least two semesters of Spanish coursework, were assessed on their ability to retain phonological elements in the short term with a Serial Non-Word Recognition (SNWR) task. Results revealed that study abroad students made greater oral gains than at-home students. However, after controlling for the effect of learning context, SNWR scores also accounted for a significant amount of variance in five of the six measures of oral proficiency gains. That is, beyond the proficiency gains attributed to learning context, students with greater phonological memory made greater gains in oral production than students with lower phonological memory in both contexts.

In another study that investigated working memory in a study abroad context, Tokowicz, Michael, and Kroll (2004) studied individual differences among students who had relatively more (15 or more months) or less (8 or fewer months) of study abroad experience. Tokowicz and colleagues proposed that learners with varying amounts of study abroad experience make different types of errors when completing a translation task from their first to second language: either nonresponse errors or meaning-based errors in which they give a related but incorrect word. Working memory was measured using an operation word task in which participants had to keep sets of words in memory while judging the accuracy of mathematical expressions. The results revealed a three-way interaction between type of error, amount of study abroad experience, and working memory capacity. Specifically, the more study abroad experience/higher working memory capacity subgroup showed a different pattern than the other groups; they had equal nonresponse and meaning-based errors rather than more nonresponse than meaning-based errors. This finding suggested that students with higher working memory could take advantage of their ability to practice a particular communicative strategy in the study abroad context.

In a study of the effects of short-term study abroad experiences, Grey, Cox, Serafini, and Sanz (2015) examined L2 lexical and grammatical development, in relation to working memory, in advanced students studying Spanish in Spain for five weeks. Students completed grammaticality judgment and lexical decision tasks at the beginning and end of their program as well as two tests of working memory: sentence span and phonological working memory. While students did improve in judging the grammaticality of Spanish sentences and making decisions about whether lexical items were actual Spanish words, these improvements were not correlated to their performance on the working memory tests.

Faretta-Stutenberg and Morgan-Short (2017) also did not find a relationship between working memory and behavioral gains for intermediate-level Spanish learners who spent 12–15 weeks abroad in Spain. They found that individual differences in procedural learning (i.e., serial response time task) predicted changes in the behavioral L2 measure of grammaticality judgment, but individual differences in working memory (i.e., operation span) predicted neurocognitive processing changes.

Another consideration proposed in the aptitude literature is whether students need a certain level or "threshold" of aptitude ability to succeed in study abroad. Sunderman and Kroll (2009) examined working memory resources in relation to learning for 48 students, 14 of whom had studied abroad for 3.8 months on average. Working memory was measured using a reading span task, and L2 measures consisted of a comprehension measure of Spanish-English translation recognition and a production measure of picture naming. Using regression analyses, they found that study abroad experience and working memory resources affected comprehension, but there was no interaction between the two factors. For production, learners with study abroad experience had greater speed and accuracy; further, the authors argue that the threshold hypothesis was supported in that without a "minimum threshold of working memory, the benefits of study abroad experience, in terms of production accuracy at the word level, are lost on these learners" (p. 93).

Considering cognitive abilities beyond working memory, Segalowitz and Freed (2004) compared students' learning during study abroad and at home with respect to individual differences in L2 lexical access and attentional control. University Spanish students completed a pretest OPI and cognitive tests at the beginning of the semester and again 13 weeks later. Participants' oral proficiency was evaluated based on four-minute excerpts of their OPIs. Overall, students in the study abroad condition showed significant gains on four of the seven oral measures, and the pretest cognitive measures of lexical access speed and efficiency were positively related to oral fluency. However, there were no interactions between learning context and cognitive abilities affecting oral gains such that these abilities were more related to gains in one learning context than in another.

Overall, the results of empirical studies are mixed regarding the role of cognitive aptitudes in students' proficiency gains from study abroad. Several studies showed a relationship between memory abilities and students' oral proficiency, particularly at lower proficiency levels. However, other aptitude abilities may play an important role in learning in the immersion context. Thus, testing a broader range of potential aptitude predictors within the same, larger, sample of students may shed light on the relative contributions of distinct cognitive constructs.

Research Question

This study addressed the following questions: Do students make significant gains in oral proficiency during their study abroad? Are there relationships between cognitive aptitudes measured by Hi-LAB and linguistic gains during study abroad? If so, what is the nature of these relationships? In order to address these questions, data were collected with university students learning Arabic on a semester-long study abroad program sponsored by an American university. Students were hypothesized to improve on their OPIs after their study abroad experience. We further explored the role of aptitude by analyzing whether their scores on cognitive and perceptual tasks predicted changes in their OPI scores.

Methods and Procedures

Participants

Over two fall semesters, 76 students of Arabic who participated in a study abroad semester in Amman, Jordan, were enrolled in the study, which was approved by the university's Institutional Review Board. All the students were L2 Arabic learners and had completed four semesters of college Arabic prior to their departure to Jordan. Three students were excluded for having prior Arabic immersion experience, and two others were excluded for not having outcome OPI scores. Therefore, 71 students (30 female, 41 male; mean age: 24 years [SD: 4.7]) were included in the final analysis. The students were experienced language learners and had studied on average two (range: 0–5) foreign languages other than Arabic; 37 students had also spent significant time abroad (three months or more) immersed in another language.

Procedures and Program

Two cohorts of students participated in the study in two consecutive years of the program. Prior to departure to Jordan, participants took the Hi-LAB (see the following) and OPI speaking pretest.[1] The postprogram OPIs were given to students at the end of their study in Jordan.

Students were in class for 2 hours a day 5 days a week over 14 weeks. The classes consisted of a review of current events and student presentations and discussion. Students were assigned to read newspaper articles for about two hours per day, some for general comprehension and some for specific information. They also spent two hours per day speaking with native speakers in Arabic outside of class, either in the community or with an assigned native speaker partner. They lived in apartments with the other students and were not asked to sign a language pledge. See Belnap et al. (2015) for more details on the study abroad program.

Materials

Oral Proficiency Interview

The American Council on the Teaching of Foreign Languages (ACTFL) OPI is a protocol for assessing spoken proficiency through an approximately 30-minute face-to-face or telephone interview conducted by certified ACTFL OPI testers. The test is based on the ACTFL proficiency scale, which includes five major levels: Novice, Intermediate, Advanced, Superior, and Distinguished. The Novice, Intermediate, and Advanced levels are divided into three sublevels: Low, Mid, and High. Because the scale is considered to be an "inverted pyramid," such that progression between adjacent levels is easier at the bottom of the scale than at the top, the proficiency levels of the ACTFL scale are ordinal in data analysis.

Language Aptitude Measures

The Hi-LAB (Doughty, 2013; Doughty, Bunting, Campbell, Bowles, & Haarman, 2007; Linck et al., 2013) consists of 10 tests of various aptitude constructs relevant for language learning: explicit learning, implicit learning, working memory, executive function, and phonological perception. The tasks that were administered are described in the following, grouped broadly by aptitude construct.

Explicit Learning

PAIRED ASSOCIATES (PA)

In this task, participants must learn 20 word pairs, each an English noun paired with a nonword. The score is the number of correctly recalled words out of 20. The test is adapted from Carroll and Sapon's (1959) test of associative memory and represents learning new vocabulary words in a foreign language.

LETTER SETS (LSET)

In this multiple-choice test adapted from Ekstrom, French, and Harman (1979), four of five sets of letters obey the same pattern, and participants select the set that violates the pattern. The score is the number of correctly chosen items out of 15, indicating the ability to consciously derive patterns and rules, such as grammar, from examples.

Implicit Learning

AVAILABLE LONG-TERM MEMORY (ALTM)

This test measures associative priming of long-term memory (Was & Woltz, 2007). There are two tasks, a priming task and a comparison task, that are interleaved throughout the test. In the priming task, participants listen to a list of five English words and are then shown two topic words, one of which is a synonym for two of the words in the list and one of which is a synonym for three of the words in the list. The participants indicate which word had more synonyms in the list. Following each list of the priming task is the comparison task, in which pairs of words are presented on the screen simultaneously, and the participant indicates whether the words have similar or different meanings. The score for the ALTM tasks depends on the rate of correct responses to primed sets of word comparisons vs. unprimed sets of words. This score may indicate a person's ability to efficiently process words with related meanings, which would help with understanding words in a foreign language.

SERIAL REACTION TIME (SRT)

In this task, adapted from Willingham, Nissen, and Bullemer (1989), four horizontally arranged boxes are shown on the screen, indicating the four positions in which an asterisk will appear. On each trial, an asterisk appears in one box and the participant must press the corresponding button on the keyboard. In blocks two, three, four, and five, the asterisks appear in a repeating pattern of length 12. This score shows how much a person adapts (speeds up) their processing of stimuli with increasing practice on the test.

Working Memory

NONWORD SPAN (NWS)

In this task, phonotactically plausible nonwords are presented. At the end of each list, participants are prompted with nonwords and must indicate whether the word was on the most recent list. The score is the number of correct answers, with a maximum possible score of 210. This test of phonological short-term memory was adapted from Gathercole, Pickering, Hall, and Peaker (2001), and shows how a person remembers information for short periods of time, even when the meaning is unclear, similar to remembering new foreign language words.

RUNNING MEMORY SPAN (RMS)

This test measures the updating subcomponent of working memory (Bunting, Cowan, & Saults, 2006; Miyake, Friedman, Emerson, Witzki, & Howerter, 2000). In this task, participants listen to lists of 12–20 letters. At the end of the list, the participants must recall the last six letters in the list, in order. The score (out of six) is the average number of letters correctly recalled in serial order per list. The score shows the ability to keep information in memory while updating that information without making mistakes, as is particularly important during interpretation from one language to another.

Executive Function

ANTISACCADE (AS)

This test was developed by Unsworth, Schrock, and Engle (2004) to measure inhibitory control (Miyake et al., 2000). In this task, visual cues are presented, indicating the location of a letter that is displayed after the cue disappears. Participants must indicate which letter is shown. In the prosaccade phase, the cue and letter appear on the same side; in the antisaccade phase, the letter appears on the opposite side of the screen. Thus, in the antisaccade phase, the participant must inhibit the tendency to look toward the cue in order to see the letter. Scoring is based on accuracy in the antisaccade phase, with higher scores indicating greater levels of ability to inhibit automatic responses, such as interference among a person's dominant and other known languages.

TASK SWITCHING (TS)

This test measures task switching (Miyake et al., 2000); participants identify whether a digit is (i) odd or even or (ii) lower or higher than five. For each set of pure blocks, participants perform the same judgment for each trial. For mixed blocks, the participants switch between performing the judgment types, based on the color of the background box. We computed a Mix Cost score, by taking the difference between the mean reaction times for mixed blocks and for pure blocks, representing the overall slowdown exhibited by the participant due to the demands of task switching. The score indicates the ability to switch between different, but related, tasks and in a foreign language context, the ability to switch between different languages. A processing speed measure is also derived from this measure by simply computing the mean reaction time in pure blocks.

Phonological Perception

PHONEMIC DISCRIMINATION OUTLIER (PDO)

There are two forms of the PDO task based on Silbert et al. (2015). For each trial, participants listen to a set of three spoken syllables; two of the sounds are examples of the same syllable, and one (the "outlier") is an example of a different syllable, which varies according to a speech sound contrast that is meaningful in a non-English language. After hearing the three sounds, participants indicate the outlier sound. In PDO 1, the two different syllables are taken from Hindi where the boundaries along the voice-onset time continuum differ from English. In PDO 2, the syllables differ according to a tonal contrast in Thai. The scores for the PDO tasks depend on the participant's accuracy in identifying the "outlier" sound in each trial and indicate the ability to hear subtle speech sound differences that do not exist in English.

Results

Proficiency Gains

We first examine the pattern of oral proficiency gains made by the students over the course of the study abroad program. On the pretest OPI, participants in the study obtained scores ranging from Novice Mid to Advanced High, but most students (62 out of the 71 in the analysis) were in the Intermediate range. The first result is that overall, there were significant proficiency gains. The pattern is displayed graphically in Figure 26.1, which gives the number of participants in each cell for each combination of pre- and posttest OPI scores in the observed ranges. The labels on the axes are abbreviations such that the first letter indicates N(ovice), I(ntermediate), or A(dvanced),

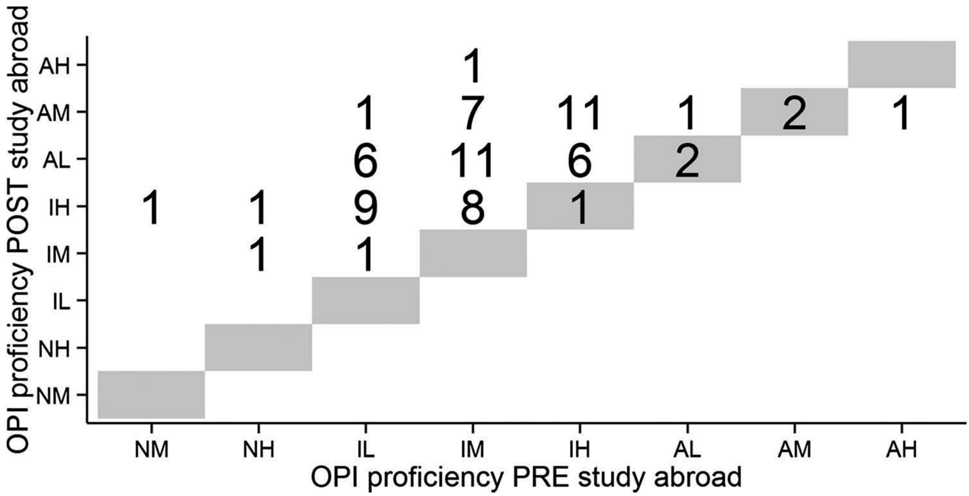

Figure 26.1 Pattern of OPI pre- and posttest scores.

while the second letter indicates L(ow), M(edium), or H(igh) within a level. The diagonal of gray-background cells represents maintenance of the same level, cells above the diagonal represent improvement, and cells below represent loss.

Clearly, the dominant pattern is gains, and the largest clusters of students are concentrated around gains of two sublevels. Only 5 out of the 71 participants (7%) maintained the same score, and only 1 lost a sublevel. Another notable pattern is that while 20 students (27%) began with an ACTFL rating of Intermediate Low or lower, every student scored above that range on the posttest. We fit an ordinal logistic mixed-effects model,[2] and the difference between pre- and posttests was significant ($\beta = 5.29$, SE = 0.69, $z = 7.63$, $p < 0.001$). These results suggest that this study abroad program is highly effective in enabling students to make substantial gains in proficiency, in the Intermediate to Advanced range of proficiency levels.

Prediction of Gains

The focus of this study is whether differences in aptitude can predict variability in oral proficiency gains on the ACTFL scale. Our approach to the issue of predicting gains using an ordinal scale was to fit models in which we predicted posttest OPI scores but included pretest OPI scores as a predictor. By adding aptitude predictors to this model, we are able to test whether aptitude contributes to prediction of final scores, above and beyond starting score and base rate of improvement. To address the issue of having 12 potential aptitude predictors with only 71 individuals' data, we applied a stepwise model selection procedure based on the corrected Akaike Information Criteria (AICc) fit statistic (Burnham & Anderson, 2002). We started with a "base" model, with no aptitude predictors, only an intercept and the pretest score as predictors. We then fit all other models that included one additional aptitude predictor (i.e., Hi-LAB test score) and compared the models using AICc. Each of the models that were better than (i.e., had a lower AICc than) the "base" model was then considered as the new "base" model, initiating the procedure again for each of these

Table 26.1 Pooled coefficients from ordinal regression for aptitude predictors

Predictor	*Coefficient estimate*	*Standard error*	*t (df)*	*p*
Running Memory Span (RMS)	.92	.30	3.08 (53.8)	.003
Paired Associates (PA)	.68	.29	2.31 (53.5)	.025
Serial Reaction Time (SRT)	−.83	.29	−2.89 (53.4)	.006
Task Switching (TS Mix Cost)	.45	.28	1.62 (53.8)	.112

better models. This recursive procedure was carried out until no additional predictors added any improvement over the current "base" model, ultimately resulting in a single "winner" model.

To address the issue that some participants were missing one or more Hi-LAB measures (eight participants were missing at least one measure due to computer or other error), we employed a multiple imputation by chained equations approach (van Buuren & Groothuis-Oudshoorn, 2011). In this approach, observed variables were used to provide (noisy) predictions for each missing data point, which preserved observed multivariate relationships in the data, thus reducing the potential for biased imputations. This was iterated to produce several complete imputed data sets, and analysis was carried out separately for each imputed data set. Finally, the results fro m each imputation were combined, according to standard algorithms (Little & Rubin, 2014). We performed this procedure using the "mice" package (van Buuren & Groothuis-Oudshoorn, 2011) in the R statistical software package (R Core Team, 2015) to produce 20 imputed data sets.[3]

We ultimately ran the model selection procedure described earlier on each of the imputed data sets. The model described was the "winner" for 19 out of the 20 imputations. We report the pooled results across imputations, as per Rubin's rules. The standard errors for the following coefficients therefore incorporate both within- and between-imputation variability. The final model coefficients for the aptitude predictors are given in Table 26.1.[4] The predictors were standardized prior to model-fitting, so the effect sizes of the coefficients are on the same (logit) scale and are thus comparable.

In this model, working memory updating (RMS) and associative memory (PA) were both implicated as significantly positive predictors, while implicit pattern learning (SRT) was a significant negative predictor. Task switching (TS Mix Cost) was positive and led to an improved model fit, but it did not reach the standard criterion for statistical significance.

Discussion

The primary goal of this study was to explore the relative contributions of a wider range of cognitive aptitudes in predicting L2 oral proficiency gains during a study abroad immersion experience, especially since prior results in the literature have been mixed. Our primary findings support working memory as the single most robust predictor of proficiency gains, consistent with previous work by O'Brien et al. (2007) and Sunderman and Kroll (2009). Interestingly, the construct of working memory updating, as measured by the RMS task, appeared more important than phonological working memory capacity, as measured by the NWS task. Whether this is a result of measurement issues or a true reflection of the specific constructs involved is a question for future research, but the result does suggest that not all measures of working memory are equal when it comes to predicting gains in oral proficiency during study abroad.

Greater ability on the RMS task, which measured the ability to continually update working memory, may have aided students in managing the real-time processing demands when interacting with native speakers. That is, students may be better able to hold in mind new words and phrases, while deciphering them and producing responses. In particular, this study abroad program placed an emphasis on interaction and required students to spent two hours per day speaking in Arabic with native speakers. Thus, an advantage in working memory may compound over time in allowing more successful and meaningful conversations that lead to learning, though more research is needed to support this interpretation. Prior research that examined advanced learners such as Grey et al. (2015) did not find a relationship with working memory capacity in the context of a short-term period abroad, and some of the same authors (Serafini & Sanz, 2016) found that the role of working memory capacity may diminish as learner proficiency increases in relation to classroom instruction. These findings could be considered in line with ours in that capacity-style span measures of working memory may not be tapping the most relevant aspects of phonological working memory for these learning environments since in the present study, our measure of working memory *updating* (RMS) was a very robust predictor, while our measure of working memory capacity (NWS) was not.

The other significant positive predictor of gains was associative or rote memory, as measured by our PA task. As this construct is typically associated more with explicit learning, we were somewhat surprised to see it emerge as a predictor in the study abroad context. Its presence suggests that at least in the context of this particular study abroad program, the ability to learn new words via conscious, rote memorization is still a strong asset, possibly because students do spend time focusing on texts in their newspaper reading class and developing vocabulary skills using apps such as Anki. The size of the effect was somewhat smaller than the effect for working memory updating, though, suggesting that it may be of secondary importance.

We also found effects for other cognitive measures that have not been previously examined in the study abroad context. A positive effect of cognitive control, as measured by the TS task, is not surprising in the study abroad context, given the exertion of cognitive control when operating in both first and second language over several months. However, the effect was fairly small and did not reach statistical significance, again suggesting that its role is likely secondary to other abilities, such as active working memory updating.

The negative effect of implicit pattern learning, as measured by the SRT task, is counterintuitive given that one of the prime benefits of study abroad is thought to be the greater amount of language input, which should provide an environment amenable to implicit learning. This negative result suggests that greater implicit learning abilities are a liability, or that lesser abilities are an asset, in the context of this study abroad program. On closer examination, the task itself appears to be measuring the ability appropriately, demonstrating the classic pattern of reaction time speed-up due to the repeated stimulus sequence as observed many times in the literature on the SRT task (Willingham et al., 1989). A replication effort for this effect would be extremely helpful as it is possible that this effect is simply a statistical anomaly, and it is highly unexpected from current theories about the role of implicit learning.

Another result from our study speaks to the question of a "threshold" effect where students need a certain level of aptitude ability to benefit from a study abroad experience, as discussed by Segalowitz and Freed (2004) and Sunderman and Kroll (2009). We performed a similar extrapolation to that used by Sunderman and Kroll (2009) to estimate a "threshold" of aptitude needed for gains.[5] The model estimated that a student roughly four standard deviations below the mean RMS score would achieve one sublevel less at posttest. Four standard deviations is an extreme outlier, less likely than 1-in-10,000. Moreover, the average gains were close to two sublevels in this program. In other words, the present model of our data predicts that only an extraordinarily low ability in working memory would be enough to reliably offset the apparent benefits of this particular program. It is possible that some kind of minimum threshold could exist that is less extreme, but our results here do not show any indication of such a threshold.

Our work suggests that in addition to other factors, such as time in country and starting proficiency, which can contribute to whether or not a student achieves proficiency gains, aptitude clearly can provide information about which students are expected to benefit more than others from the experience. We explored a wide range of aptitude constructs with a large sample size and found that working memory updating, rote memory, and task-switching abilities were positively related to proficiency gains; in contrast, implicit learning skills showed a negative relationship. More research is needed to understand the processes by which these abilities may influence learning in the immersion context, but our work represents a first step in examining the relative contributions of a variety of cognitive abilities.

Implications: Recommendations for Practice

Given these results, we draw the following implications from this study. First, the fact that we found an impact of individual differences in cognitive aptitude on gains in study abroad contexts suggests that some students are better cognitively equipped to make greater gains during a study abroad experience than others, all else being equal. Thus, testing students' cognitive aptitudes prior to going abroad can be an effective practice, the results of which can be used to (i) inform student selection to competitive study abroad programs, (ii) tailor study abroad experiences for learners, and (iii) adjust student behaviors while abroad.

Second, while aptitude does appear to make a difference, we observed substantial proficiency gains virtually across the board, even for those with the lowest aptitude. This suggests that while aptitude information may be useful in projecting who may make the greatest gains, in a good program that provides immersion in the target language and culture coupled with intensive classroom study and out-of-class immersive activities, even students with very low aptitude may still make significant gains.

Third, while our results most strongly implicate aspects of memory, both working memory and associative rote memory, how these constructs are measured may matter in terms of how predictive they are. More generally, if students are to be tested for aptitude, the measurement matters. In particular, it may be the case that measures of working memory updating are better predictors than span tasks that measure working memory capacity, though this is an area that needs further investigation.

Taking these findings together, we suggest that implementing educational interventions that are matched to students' aptitudes prior to studying abroad may further improve their outcomes. The strongest predictor of gains out of the aptitude measures was the updating component of working memory, measured using the RMS task; this strong result makes it a good target for interventions in order to maximize impact. Thus, one possibility is that targeted production practice for students with lower working memory, via text chat, conversation groups, or other means, prior to studying abroad may support their speaking gains while abroad. Another possibility might be that students with lower working memory could be given alternate strategies for interacting with native speakers while abroad, such as trying to discuss more familiar topics while building speaking skills, in order to alleviate the task-based demands on working memory. In this particular program, students could be assigned to speaking partners who can adjust their interactions based on students' proficiency and aptitude profiles (Bown, Dewey, & Belnap, 2015).

Limitations/Future Directions

One limitation of the present study is that it examines aptitude only in relation to global proficiency measures, which may not fully capture students' language gains. Future studies should consider how, in addition to global proficiency measures, analyses relative to more fine-grained measures, such as fluency and complexity, can inform which aptitudes are relevant for study abroad. In a preliminary analysis of these students' semantic fluency and spoken production, Lancaster et al. (2016) have examined how fine-grained measures related to the Hi-LAB tests. Another limitation of this study is that we did not use the aptitude information to tailor the study abroad experience for learners. In the future, we wish to use these profiles to inform opportunities for tailoring study abroad experiences based on cognitive aptitudes. In addition, we propose that aptitude-by-treatment interaction (ATI; see Vatz, Tare, Jackson, & Doughty, 2013 for a review) studies be conducted to match different aptitude profiles to different study abroad approaches. As a result of this research, learners would be able to use aptitude information diagnostically and find out the best strategies to use in order to optimize learning, which may, in turn, improve linguistic gains during study abroad.

Key Terms

Study abroad
Immersion
Aptitude
Hi-LAB
Working memory
Explicit learning
Implicit learning
Executive function
Oral Proficiency Interview (OPI)
Speaking proficiency
ACTFL
Arabic

Notes

1 Hi-LAB testing was conducted using an online research platform; some students completed these tasks in a proctored session and others completed it self-proctored, depending on external factors. Oral Proficiency Interviews were completed by phone for the students tested in the first year and in-person for the students tested in the second year.

2 A random intercept by participant was included in this model, to reflect the repeated-measures nature of this analysis, in order to model the fact that participants started at a

range of levels. A random slope of time by participant was not able to be fit with this data in this kind of model, given that there is only one data point for each combination of participant by time. This analysis uses the number of sublevels gained as an ordinal dependent variable. We argue that this measure is not sufficient for the full aptitude analysis, but it suffices here, since the pattern of gains is so clear, even without a statistical test.

3 It is more typical to use a smaller number of imputations, such as five, but a larger number was used here in order to provide more stable results.

4 Due to space constraints, the parameters for the ordinal cut-points and the ordinal pre-test scores are not included. The effects of pre-test scores were, as expected, significant and generally increase from lower to higher levels.

5 In the final ordinal regression model, the average logit "distance" between the estimated cut-points in the dependent measure is four times the coefficient of the RMS predictor.

Further Reading

Belnap, R. K., Bown, J., Dewey, D. P., Belnap, L., & Steffen, P. (2016). Project perseverance: Helping students become self-regulating learners. In T. Gregersen, P. MacIntyre, & S. Mercer (Eds.), *Positive psychology in SLA* (pp. 282–301). Bristol: Multilingual Matters. (This chapter provides an overview of Brigham Young University's intensive Arabic semester abroad program, including quantitative results of student learning outcomes and their evaluation of various interventions.)

Bown, J., Dewey, D. P., & Belnap, R. K. (2015). Student interactions during study abroad in Jordan. In R. Mitchell, N. Tracy-Ventura, & K. McManus (Eds.), *Social interaction, identity and language learning during residence abroad* (pp. 199–222). Eurosla Monographs Series, 4. (This chapter reports on a qualitative study of the speaking experience of 82 students of Arabic and factors that influenced the quality of their interactions, documenting in particular the value of regular appointments with trained tutors.)

Linck, J. A., Hughes, M. M., Campbell, S. G., Silbert, N. H., Tare, M., Jackson, S. R., … Doughty, C. J. (2013). Hi-LAB: A new measure of aptitude for high-level language proficiency. *Language Learning, 63*, 530–566. (This study examines which cognitive aptitudes can discriminate between average- and high-level language learners.)

Vatz, K., Tare, M., Jackson, S., & Doughty, C. (2013). Aptitude-treatment interaction studies in second language acquisition: Findings and methodology. In G. Granena & M. H. Long (Eds.), *Sensitive periods, language aptitude, and ultimate L2 attainment* (pp. 273–292). Amsterdam, the Netherlands: John Benjamin's Publishing. (This chapter outlines the state of the science for evidence of interactions between cognitive aptitudes and language-training techniques as well as evaluates the strengths and weaknesses of the literature.)

References

Belnap, R. K., Bown, J., Dean, E. M., Dewey, D. P., Schouten, L. J., Smith, A. K., Smith, R. K., & Taylor, J. R. (2015). Project perseverance and study abroad. *Al-Arabiyya, 48*, 1–21.

Bown, J., Dewey, D. P., and Belnap, R. K. (2015). Student interactions during study abroad in Jordan. In R. Mitchell, N. Tracy-Ventura, & K. McManus (Eds.), *Social interaction, identity and language learning during residence abroad*, Eurosla Monographs Series, *4*, 199–222.

Brecht, R., Davidson, D., & Ginsberg, R. (1993). *Predictors of foreign language gain during study abroad.* Washington, DC: National Foreign Language Center.

Bunting, M. F., Cowan, N., & Saults, J. S. (2006). How does running memory span work? *The Quarterly Journal of Experimental Psychology, 59*, 1691–1700.

Burnham, K. P., & Anderson, D. R. (2002). *Model selection and multimodel inference*. New York, NY: Springer.

Carroll, J. B., & Sapon, S. M. (1959). *Modern language aptitude test*. San Antonio, TX: Psychological Corporation.

Davidson, D. (2010). Study abroad: When, how long, and with what results? New data from the Russian front. *Foreign Language Annals, 43*(1), 6–26.

Dewey, D. (2008). Japanese vocabulary acquisition by learners in three contexts. *Frontiers: The Interdisciplinary Journal of Study Abroad, 15*, 127–273.

Doughty, C. J. (2013). Optimizing post-critical-period language learning. In G. Granena & M. H. Long (Eds.), *Sensitive periods, language aptitude, and ultimate L2 attainment* (pp. 153–174). Oxford, UK: Wiley-Blackwell.

Doughty, C. J., Bunting, M. F., Campbell, S. G., Bowles, A. R., & Haarmann, H. (2007). *The development of the High-level Language Aptitude Battery*. Technical Report: Center for Advanced Study of Language. University of Maryland, College Park.

Ekstrom, R. B., French, J. W., & Harman, H. H. (1979). Cognitive factors: Their identification and replication. *Multivariate Behavioral Research Monographs, 79*(2), 3–84.

Faretta-Stutenberg, M., & Morgan-Short, K. (2018). The interplay of individual differences and context of learning in behavioral and neurocognitive second language development. *Second Language Research*, *34*, 67–101.

Foster, P. (2009). Lexical diversity and native-like selection: The bonus of studying abroad. In B. Richards, M. Daler, D. Malvern, P. Meara, J. Milton, & J. Treffers-Daller (Eds.), *Vocabulary studies in first and second language acquisition* (pp. 91–106). Basingstoke, UK: Palgrave Macmillan.

Freed, B. (1995). *Second language acquisition in a study abroad context*. Philadelphia, PA: John Benjamins.

Gathercole, S. E., Pickering, S. J., Hall, M., & Peaker, S. M. (2001). Dissociable lexical and phonological influences on serial recognition and serial recall. *The Quarterly Journal of Experimental Psychology, 54A*, 1–30.

Grey, S., Cox, J., Serafini, E. J., & Sanz, C. (2015). The role of individual differences in the study abroad context: Cognitive capacity and language development during short-term intensive language exposure. *Modern Language Journal, 99*(1), 137–157.

Lancaster, A. K., Tare, M., Bonilla, C., Golonka, E., Mermel, M., & Jackson, S. R. (2016, September). *Spoken fluency of Arabic second-language learners during study abroad.* Paper presented at the conference of Investigating Arabic Second Language Learning: Empirical Findings and Trends, Ann Arbor, MI.

Larson-Hall, J., & Dewey, D. (2012). An examination of the effects of input, aptitude, and motivation on the language proficiency of missionaries learning Japanese as a second language. In L. Hansen (Ed.), *Second language acquisition abroad: The LDS missionary experience* (pp. 45–51). Amsterdam, the Netherlands: John Benjamins.

Linck, J. A., Hughes, M. M., Campbell, S. G., Silbert, N. H., Tare, M., Jackson, S. R., … Doughty, C. J. (2013). Hi-LAB: A new measure of aptitude for high-level language proficiency. *Language Learning, 63*, 530–566.

Linck, J. A., Osthus, P., Koeth, J. T., & Bunting, M. F. (2014). Working memory and second language comprehension and production: A meta-analysis. *Psychonomics Bulletin Review, 21*, 861–883.

Little, R. J., & Rubin, D. B. (2014). *Statistical analysis with missing data*. New York, NY: John Wiley & Sons.

Llanes, A. (2011). The many faces of study abroad: An update on the research on L2 gains emerged during a study abroad experience. *International Journal of Multilingualism, 8*, 189–215.

Meara, P. M. (2005). *Llama language aptitude tests*. Swansea, UK: Lognostics.

Miyake, A., Friedman, N. P., Emerson, M. J., Witzki, A. H., & Howerter, A. (2000). The unity and diversity of executive functions and their contributions to complex "frontal lobe" tasks: A latent variable analysis. *Cognitive Psychology, 41*, 49–100.

O'Brien, I., Segalowitz, N., Freed, B., & Collentine, J. (2007). Phonological memory predicts second language oral fluency gains in adults. *Studies in Second Language Acquisition, 29*, 557–582.

R Core Team. (2015). *R: A language and environment for statistical computing.* Vienna, Austria: R Foundation for Statistical Computing.

Segalowitz, N., & Freed, B. F. (2004). Context, contact, and cognition in oral fluency acquisition. *Studies in Second Language Acquisition, 26*, 173–199.

Serafini, E. J., & Sanz, C. (2016). Evidence for the decreasing impact of cognitive ability on second language development as proficiency increases. *Studies in Second Language Acquisition, 38*(4), 607–646.

Silbert, N., Smith, B. K., Jackson, S. R., Campbell, S. G., Hughes, M. M., & Tare, M. (2015). Non-native phonemic discrimination, phonological short term memory and word learning. *Journal of Phonetics, 50*, 99–119.
Sunderman, G., & Kroll, J. (2009). When study-abroad experience fails to deliver: The internal resources threshold effect. *Applied Psycholinguistics, 30*(1), 79–99.
Tokowicz, N., Michael, E., & Kroll, J. (2004). The roles of study-abroad experience and working-memory capacity in the types of errors made during translation. *Bilingualism: Language and Cognition, 7*, 255–272.
Unsworth, N., Schrock, J. C., & Engle, R. W. (2004). Working memory capacity and the anti-saccade task: Individual differences in voluntary saccade control. *Journal of Experimental Psychology: Learning, Memory, & Cognition, 30*, 1302–1321.
van Buuren, S., & Groothuis-Oudshoorn, K. (2011). Mice: Multivariate imputation by chained equations in R. *Journal of Statistical Software, 45*(3), 1–67.
Vatz, K., Tare, M., Jackson, S., & Doughty, C. (2013). Aptitude-treatment interaction studies in second language acquisition: Findings and methodology. In G. Granena & M. Long (Eds.). *Sensitive periods, language aptitude, and ultimate L2 attainment* (pp.273–292). Amsterdam, the Netherlands: John Benjamins Publishing Company.
Was, C. A., & Woltz, D. J. (2007). Reexamining the relationship between working memory and comprehension: The role of available long-term memory. *Journal of Memory and Language*, *56*, 86–102.
Willingham, D. B., Nissen, M. J., & Bullemer, P. (1989). On the development of procedural knowledge. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 15*, 1047–1060.