

Relationships of Attitudes Toward Homework and Time Spent on Homework to Course Outcomes: The Case of Foreign Language Learning

Charles B. Chang
Rice University

Daniel Wall
Columbia University

Medha Tare and Ewa Golonka
University of Maryland, College Park

Karen Vatz
University of Vermont

In previous studies of homework in core academic subjects, positive student attitudes toward homework were linked to higher achievement, whereas time spent on homework showed an inconsistent relationship with achievement. This study examined the generalizability of these findings to foreign language learning by analyzing 2,342 adult students' attitudes toward assigned homework, time spent on assigned homework, and achievement outcomes in a variety of foreign language courses. Student ratings of the relevance of homework, the usefulness of feedback provided on homework, and the fairness of homework grading were positively correlated with teacher-assigned grades and standardized proficiency test scores in listening, reading, and speaking. Reported time spent on homework, however, was negatively correlated with these measures. In hierarchical regression analyses, all homework-related variables emerged as significant predictors of outcomes after controlling for potential covariates such as language learning aptitude, demographic variables, and affective factors. Thus, these results provide evidence that language course outcomes are positively associated with attitudes toward homework but negatively associated with time spent on homework. Possible interpretations of these findings are discussed. We suggest that the negative association follows in part from the opportunity cost of time spent on assigned homework, which decreases time spent on individualized study that may be more beneficial for improving language course outcomes.

Keywords: assigned homework, attitudes, time spent, grade point average, standardized test scores

Homework is nearly ubiquitous in educational settings, yet its effectiveness as a pedagogical tool is questioned to this day. In this article, we report results of the first study to examine the relationship of homework-related variables to achievement in foreign language (FL) courses. In doing so, this investigation responds to a problematic bias in education research toward examining

achievement in core academic subjects (e.g., reading, math, science) to the exclusion of FLs. Although there is some evidence that homework is related positively to achievement in core subjects, FL achievement may not be influenced by homework in the same way because of two unique characteristics of language acquisition: a high degree of implicit learning (Eckman, Iverson, Fox, Jacewicz, & Lee, 2011; Williams, 2009) and social and interactional requirements (Burling, 1981; Gass, Mackey, & Pica, 1998; Kuhl, Tsao, & Liu, 2003; Long, 1981; Schumann, 1986; Seliger, 1977; Watanabe & Swain, 2007; Wells, 1981). Unlike that of other subjects, knowledge of language is largely unconscious; is routinely acquired without explicit instruction; and is inherently social, in that rules of conversation and cultural norms are fundamentally based in social interaction. Consequently, although traditional homework exercises done in isolation may aid in the learning of algebra, for example, it remains unclear whether this type of explicit, noninteractional reinforcement provides the same benefits in the learning of an FL. As yet, no study in the literature on homework has demonstrated how homework relates to achievement in learning an FL. This gap in the literature is addressed in the current study.

To examine the relationship between homework-related variables and FL course outcomes, we conducted a large-scale analysis of survey and outcome data from adult students in an intensive FL program. Adult students are of particular interest with respect to FL learning in the United States because, even though language

This article was published Online First April 7, 2014.

Charles B. Chang, Department of Linguistics, Rice University; Daniel Wall, Center for Decision Sciences, Columbia University; Medha Tare and Ewa Golonka, Center for Advanced Study of Language, University of Maryland, College Park; Karen Vatz, Department of Romance Languages and Linguistics, University of Vermont.

This material is based upon work supported, in whole or in part, with funding from the United States Government; however, any opinions, findings and conclusions, or recommendations expressed here are the authors' and do not necessarily reflect the views of any of the authors' institutions or any agency of the United States Government. We are grateful to audiences at the University of Maryland for helpful feedback; to Shawn Maloney, Kathy Rhoad, and Rachel Strong for assistance with parts of the background research; and to the language school for permitting us to publish these data.

Correspondence concerning this article should be addressed to Charles B. Chang, Rice University, Department of Linguistics, P.O. Box 1892, Houston, TX 77251. E-mail: cbchang@post.harvard.edu

acquisition is typically less successful following a “critical period” in childhood (Johnson & Newport, 1989; Patkowski, 1990), Americans still tend to begin FL study late in life. Their FL study begins well after this critical period (Pufahl, Rhodes, & Christian, 2000, 2001; Rhodes & Pufahl, 2009), and few manage to attain a level of professional working proficiency in an FL during high school. Because various professions from government to business now call for FL skills (Berdan, 2012; Byrne, 2013; Davidson, 2012; Gragreen, 2013; Helmore, 2013; Skorton & Altschuler, 2012), this often leads to (re-)learning of an FL in adulthood. For this reason, the findings of this study are especially relevant for FL educators in the United States, many of whom are working with adults instead of children. By examining adult learners, the present study also contributes much-needed data on continuing education to the current literature on homework, which is limited largely to students at the primary and secondary school levels. Note, however, that this study pertains only to assigned homework (i.e., homework exercises assigned by the teacher)—not to “homework” in the general sense of the word (i.e., any learning work completed at home or outside the classroom)—because assigned homework is what participants were asked about on the survey they completed.

The article is organized as follows. In the next section, we review the published research on homework, concentrating in particular on studies that shed light on how achievement is influenced by attitudes toward homework and time spent on homework. In the following sections, we describe our study in more detail and present the results of an extensive series of analyses investigating the relationship between homework-related variables and FL course outcomes. Finally, in the last two sections, we discuss possible interpretations of the results and implications for FL education.

Background

Attitudes Toward Homework

Relatively little research has investigated student attitudes toward homework in the context of FL learning or in relation to achievement; however, attitudes toward homework in other subjects have been shown to be related to achievement measures in at least one study (Cooper, Lindsay, Nye, & Greathouse, 1998). For this study, 709 students in Grades 2–12 completed the Homework Process Inventory, a questionnaire designed to gauge attitudes regarding homework and to determine whether these attitudes could be linked to student achievement as measured by teacher-assigned grades and state standardized testing (which included reading, math, science, and social studies scores). Correlational analyses showed that although the homework attitudes of lower grade students (Grade 4 and below) were not significantly associated with achievement, the homework attitudes of upper grade students (Grade 6 and above) were positively associated with achievement. Upper grade students’ attitudes were positively correlated with their reports of the portion of homework completed and the composite measure of their time spent on homework, as well as with their grades (but not with their test scores). The size of the correlation between upper grade students’ attitudes and grades was quite small ($r = .11$). Nevertheless, these results led Cooper et al. (1998) to suggest that attitudes toward homework

may play a more important role in achievement as students mature and perform more self-regulated learning.

A motivational tactic used to improve attitudes toward homework is to provide students with a choice of homework exercises, and one study of high school students has suggested that such provision of choice increases motivation and performance outcomes and is important for supporting autonomy and intrinsic motivation of students in the classroom (Patall, Cooper, & Wynn, 2010). The 207 students in this study were randomly assigned to either a homework-choice condition or a no-homework-choice condition; students in the homework-choice condition were given two options for their homework exercise (similar in content and difficulty level) and asked to decide which homework exercise to complete, and students in the no-homework-choice condition were secretly paired with those in the homework-choice condition (such that the exercise chosen by the student in the homework-choice condition was assigned to the corresponding student in the no-homework-choice condition). All students participated in the homework-choice condition during one of two units of instruction, to allow the effects of choice to be analyzed within an individual student, and their motivation and mastery of the material were assessed at the end of each unit. Results indicated that, compared to students in the no-homework-choice condition, students in the homework-choice condition completed more homework, scored better on unit tests, felt more competent about the homework, and reported more interest and enjoyment in doing the homework. These findings are in accord with self-determination theory (Deci & Ryan, 1985; Ryan & Deci, 2000)—which states that choice is central to supporting feelings of autonomy, motivation, and healthful functioning—and consistent with the findings of a meta-analysis of 41 studies examining the effect of choice on intrinsic motivation and related outcomes (Patall, Cooper, & Robinson, 2008). Thus, studies of homework choice further support the notion that attitudes toward homework are positively related to achievement outcomes.

In short, the literature on attitudes toward homework provides evidence that positive attitudes are associated with higher achievement, at least on some metrics. That positive attitudes toward homework are linked with higher achievement makes intuitive sense; after all, however homework may benefit students’ learning, those who are more inclined to think positively of homework are probably better positioned to extract benefits from it. The positive link between achievement and attitudes is, therefore, unsurprising. In contrast, the relationship between achievement and time spent on homework is much more controversial, and this is the topic to which we turn next.

Time Spent on Homework

Although there is, to our knowledge, no published research on the efficacy of homework in the context of adult FL learning, findings on primary and secondary school populations outside the area of FL are abundant, albeit mixed. The mixed nature of empirical findings is evident in a meta-analysis of studies relating time spent on homework to subsequent achievement in multiple, primarily nonlanguage subjects in Grades K–12 (Cooper, Robinson, & Patall, 2006). Out of 69 correlations reviewed in this analysis, 50 showed that increased time spent on homework was associated with increased achievement, and 19 showed that in-

creased time was associated with decreased achievement. In addition, the analysis showed that within the positive results, the positive relationship was much stronger for secondary students (students in Grades 7–12) than for elementary school students, suggesting that older students differ from younger students in the benefits they derive from homework.

One of the more extensive studies reporting a positive association between time spent and outcomes is Keith and Cool's (1992) structural equation modeling study of academic achievement in high school students. This study examined the effects of several predictors—intellectual ability, quality of instruction, academic motivation, quantity of academic coursework, and time spent on homework—on a composite measure of achievement that included reading, math, science, writing, and civics standardized test scores. Results showed strong positive effects of intellectual ability and coursework ($r = .74$ and $r = .62$, respectively), as well as a weak positive effect of time spent on homework ($r = .30$). These findings led to the conclusion that “homework has a small but meaningful effect on high school seniors’ achievement” (Keith & Cool, 1992, p. 215).

Other studies, however, have noted complications in examining the effects of time spent on homework. In two studies of seventh and ninth graders, Smith (1990, 1992) examined how student reports of time use, including time spent on homework, related to achievement outcomes for reading, language, math, and overall achievement. In the first study, Smith (1990) did not find a significant relationship between reported time and achievement outcomes but noted that the results were potentially confounded by intellectual ability. Consequently, Smith (1992) followed up on his initial null results by examining the same participants 2 years later; earlier achievement scores were included as a control for ability. In this study, too, no relationship was found between time spent on homework and academic achievement as measured by a combination of reading, language, and math scores. These findings thus provide increased confidence in the original null results.

Although conflicting results across studies such as Keith and Cool (1992) and Smith (1992) complicate the interpretation of research examining time spent on homework, what makes this literature even more complex is the fact that conflicting results are reported even within the same study, such as the correlational study of Cooper et al. (1998). Their results showed that time spent was not significantly correlated with standardized test scores but was significantly correlated with teacher-assigned grades. The direction of this relationship, however, was the opposite for lower and upper grade students: Whereas time spent by lower grade students was negatively correlated with their grades ($r = -.19$), time spent by upper grade students was positively correlated with their grades ($r = .17$). This discrepancy between age groups was never explained; however, it suggests that spending more time on homework at lower grades is not as helpful as spending more time on homework at upper grades, perhaps because the type of homework assigned at upper grades requires a larger time investment by nature.

As alluded to by Smith (1990, 1992), an important consideration in interpreting findings on time spent is the presence of covariates. This point is made clear in a study of math homework variables and math achievement in ninth graders from the Netherlands (de Jong, Westerhof, & Creemers, 2000). Using scores from math ability tests as an index of prior math knowledge, this study

showed a negative correlation between prior knowledge and time spent on homework, as well as a negative correlation between time spent on homework and achievement. The latter correlation, however, likely resulted from the negative link between prior knowledge and time spent, as the time reportedly spent on homework by students was “strongly influenced by their prior knowledge of mathematics” (de Jong et al., 2000, p. 147). Given the same homework, students who scored higher on the test of prior knowledge reported spending approximately nine minutes less on homework per lesson than students who scored lower (a large difference considering that the average reported time spent on homework per lesson was 30 minutes).

Thus, although the literature examining time spent on homework in relation to achievement shows a trend toward positive association, on the whole the findings of this literature are mixed. The disparity across studies is likely due at least in part to methodological diversity (e.g., in subject areas and age groups investigated). For example, benefits of homework reinforcement may differ depending on the degree to which the given subject lends itself to being learned without interaction or the degree to which the given learner population is prepared to extract benefits from this kind of reinforcement. We consider the implications of these types of potential variation for predictions in our study below.

Research Questions and Predictions

Given the paucity of research on the efficacy of FL homework, the current study examined the relationship of time spent on homework and attitudes toward certain aspects of homework to achievement outcomes in a range of FL courses. Whereas the majority of the aforementioned literature focused on younger, preuniversity populations, our study population comprised young adults who were learning an FL for career purposes, as is becoming increasingly common in the United States. Thus, it is important to note that our study population may differ from the populations investigated in previous work in gains from homework, and additional factors not relevant to younger students (e.g., higher educational attainment, professional motivations) may affect the relationships explored here.

This study was aimed at investigating two main research questions related to FL homework:

1. How are attitudes toward homework and time spent on homework related overall to FL course outcomes?
2. Once background variables are controlled, do homework-related variables still show a significant relationship to FL course outcomes?

To address these questions, we analyzed students’ evaluations of homework and their learning outcomes (teacher-assigned grades and standardized test scores) in several different FLs. The construct of *time spent* was measured in terms of one overall assessment, whereas the construct of *attitudes toward homework* was measured in terms of three subconstructs: perceived relevance, perceived usefulness of feedback, and perceived fairness of grading. Breaking the construct of attitudes into these subconstructs was done for two reasons. First, compared to the larger construct, these subconstructs were expected to be evaluated in a more consistent manner across participants because of less between-

participant variation in the weighting of various subconstructs. Second, although we did not expect different subconstructs of attitudes to show different relationships with outcomes, we considered it prudent to allow for this possibility. Thus, compared to previous studies, this study carried out a more fine-grained analysis of attitudes vis-à-vis outcomes.

In regard to Question 1, we predicted that FL course outcomes, like outcomes in other subjects, would show a positive relationship with attitudes toward homework but a more complex relationship with time spent on homework. In particular, we did not expect to see a positive relationship between time spent on homework and FL course outcomes due to the different nature of language learning in comparison to subjects such as reading and math. Language learning is widely regarded as a process that involves a great deal of implicit learning as well as social and interactional requirements. Language knowledge can be acquired naturalistically, without formal instruction, as attested by immigrants to the United States who learn English through their social interactions without ever setting foot in a classroom. On the other hand, even the most motivated student would be hard-pressed to attain native-like proficiency in an FL without interaction with native speakers. These facts suggest that mastery of an FL may not follow from traditional homework exercises done in isolation in the same way as mastery of other subject areas does. On the contrary, FL students who spend all of their out-of-class time completing traditional, prescribed homework exercises may end up underperforming relative to those who spend that time engaging in other, more tailored activities they find more helpful for improving their FL proficiency (e.g., speaking practice with a conversation partner). This kind of effect could thus result in a negative relationship between time spent on homework and FL course outcomes.

In regard to Question 2, we predicted that the relationships we observed in correlational analyses abstracting away from the influence of other variables would hold up after controlling for five background factors that have been linked to language learning outcomes: language aptitude, age, sex, education, and motivation. These background factors were not the focus of our study, but in general we expected their effects to be consistent with patterns previously reported in the language learning literature. Thus, we expected outcomes to show a positive relationship with language aptitude, education, and motivation (Flege, 1987; Lett & O'Mara, 1990; Schumann, 1986; Stølen, 1987) but a negative relationship with age (Flege, 1991; Flege, Munro, & Mackay, 1995; Flege, Yeni-Komshian, & Liu, 1999; Huang, 2013; Wode, 1994). With regard to sex, however, the literature was ambiguous and did not lead us to expect a specific kind of effect. For instance, the female advantage for language that has been found in first language acquisition (Berninger, Nielsen, Abbott, Wijsman, & Raskind, 2008; Huttenlocher, Haight, Bryk, Seltzer, & Lyons, 1991; Kimura, 1999) often fails to be evident in second language acquisition (Brantmeier, Schueller, Wilde, & Kinginger, 2007; Hyde & Linn, 1988); moreover, when an effect of sex is evident, the effect is not consistent across studies, sometimes favoring females and sometimes favoring males (Oxford, 1993; Tercanlioglu, 2004), and it is often unclear to what extent the effect is independent of affective differences between female and male learners of a second language (Gardner & Lambert, 1972; Kobayashi, 2002).

Because our predictions followed from general logic that applied to all languages, we did not expect any observed relation-

ships between aspects of homework and outcomes to differ radically across languages. Nevertheless, we considered the possibility that such relationships could be modulated by the degree of difficulty of the language being acquired. It is conceivable, for example, that relatively difficult languages might by nature require more attitudinal and temporal resources to be acquired successfully than relatively easy languages would; in this case, positive attitudes toward homework and time spent on homework might show a stronger relationship to outcomes in more difficult languages than in less difficult languages. Thus, in our correlational analyses we examined not only overall patterns but also patterns within each language, and in our regression analyses we included the factor of language in all of our models.

In both the analyses addressing Question 1 and those addressing Question 2, we expected effects of homework-related variables to be significant but small, as previous studies (e.g., Cooper et al., 1998; Lett & O'Mara, 1990) have consistently reported modest effect sizes of homework-related variables and other predictors of FL proficiency. These modest effect sizes suggest, on the one hand, that there are many factors influencing FL proficiency in a classroom setting that are difficult to control (e.g., teacher-student dynamic, nature of instruction) and, on the other hand, that the influence of factors that are straightforward to control tends to be relatively small in comparison. Thus, although we predicted that effects of attitudes toward homework and time spent on homework in the current study would be significant, we also predicted that these effects would be small, consistent with the magnitude of effects previously reported in the literature.

Method

Participants and Learning Context

Participants were 2,342 students taking an intensive FL course who responded to a course-end survey during the time period of 2010–2011. They were studying an FL for career purposes in a U.S.-based language program and had no prior knowledge of the language they were studying. All were over the age of 18, spoke English as their native language, and did not speak any other language natively. They were drawn from 163 different classes in the program, with an average of 14 students coming from the same class.

Prior to beginning their language course, participants completed additional surveys that provided information about their level of education, their level of motivation for learning the given FL, and their language learning aptitude. Participants reported their level of education as one of nine ordered response options (ranging from *did not complete high school* to *doctorate*) and their level of motivation from among five ordered response options (ranging from *I would prefer to do something else rather than study a foreign language* to *assigned language is my first choice*). Language learning aptitude was measured by performance on the Defense Language Aptitude Battery (DLAB), a standardized test of language aptitude comprising multiple-choice, inductive reasoning items that quantify an individual's ability to accomplish cognitive tasks related to language learning (Petersen & Al-Haik, 1976). DLAB scores are a highly reliable measure of language aptitude (overall reliability of .90; Petersen & Al-Haik 1976, p. 378) and also a strong predictor of language course outcomes,

outperforming many other predictors (including alternative aptitude tests) with a correlation between overall score and course grade of .43 (Peterson & Al-Haik, 1976, p. 376). Previously administered via paper and pencil, the DLAB is now computer based (Rodgers, 2013).

Participants' courses were in seven languages (A–G)¹ representing four different language families and a wide range of difficulty, from Group I (the easiest for native English speakers to acquire) to Group IV (the hardest for native English speakers to acquire; for more information on the classification of foreign languages in terms of difficulty, see the difficulty groups reported by the Foreign Service Institute, 1973, and the National Virtual Translation Center, 2007). The goal of each language course was to get students to the level of “limited working proficiency” on the Interagency Language Roundtable (ILR) scale (Herzog, n.d.). In each course, teachers followed a standardized curriculum with the same core homework, which they supplemented with original exercises of their own design. That is to say, there was limited variability in the quantity of homework assigned by teachers of a given course.

The assigned homework in the courses was mandatory, and its completion was generally monitored, as homework was often reviewed in class and graded. The homework materials in all seven languages were typical FL learning materials consisting of course workbooks and supplementary assignments designed by the teachers. The workbook exercises required a great deal of reading, listening, and writing in the target language, with fewer opportunities to practice speaking. The activities often involved translation, listening to or reading a target language passage, and responding to questions to indicate comprehension. Homework grading varied by teacher, but in all classes homework grades constituted a small percentage of students' final grades that was prescribed by the language program.

Survey Data

Students in the language program completed a course-end survey about various topics related to their language course. Their completion of this survey was effectively anonymous, as their responses could not be linked back to them by anyone in the language program. Therefore, it is reasonable to believe that students' responses on this survey represented an honest assessment of their homework behaviors and of their language course.

The data subjected to analysis comprised the subset of five survey items that elicited responses relevant to homework (see Table 1). Data were prepared for statistical analysis by coding responses as follows. Responses about relevance were coded as 0 (“not at all relevant”) to 2 (“very relevant”); the “cannot determine” option did not occur in the data set. Responses about feedback and grading were coded as 0 (“strongly disagree”) to 3 (“strongly agree”); the “don't know, no opinion” option did not occur in the data set. Responses about time were coded as 0 (“none”) to 6 (“more than 2.5 hr”). In addition, qualitative comments regarding homework provided in response to the free-response question were examined for recurring themes.

Free-response comments, which were provided by about one fifth of students, were analyzed in the following manner. First, three researchers each read through all of the comments to get a general sense of the data set and identify the major themes regard-

ing homework. One researcher then read through a third of the comments again to create a set of preliminary codes. Next, the other two researchers read through the remaining comments to determine how well these codes fit the major themes they discovered. The three researchers met to discuss their findings and refined the set of preliminary codes, resulting in 20 final codes (e.g., “homework is busywork”; “homework leaves no time for self-study”). Finally, the three researchers independently read through their assigned comments again to apply the final codes. The number of times each code occurred in the data set was tallied, allowing us to identify the themes that were most consistently and frequently found across language groups.

In sum, the data on homework comprised quantitative data on four homework variables: relevance of assigned homework to course content (HWRelevance), usefulness of feedback provided on assigned homework (HWFeedback), fairness of the grading of assigned homework (HWGrading), and time spent on assigned homework (HWTime). In addition to examining these quantitative data, we examined qualitative data about homework, primarily in regard to feedback and grading.

Outcome Data

Outcome data comprised two measures of students' mastery of their respective FL: a final grade point average (GPA), based on teacher-assigned grades in six ordered sections of the course, and standardized proficiency test scores from the end of the course. GPA was on a 4.0 scale and included tests, participation, and homework. The proficiency tests examined were the Defense Language Proficiency Test (DLPT) and the Oral Proficiency Interview (OPI), such that each student had a total of three scores: DLPT scores for the subcategories of listening (DLPT-L) and reading (DLPT-R) and an OPI score for speaking. These scores followed the ILR scale of 0–5 (Clark & Clifford, 1988; Herzog, n.d.).

Originally developed by the Department of Defense, the DLPT is a computer-administered “battery of tests used to assess native English speakers' reading and listening skills in a wide range of foreign languages” (Miles, 2005). Lasting six hours, the test is “designed to measure proficiency in the target language regardless of how it has been acquired”; as such, “its content is not tied to any particular language-training program” (Defense Language Institute Foreign Language Center, 2009, p. 5). Test content includes authentic and real-life source materials (e.g., signs, newspapers, radio broadcasts); covers a wide range of topics such as society, culture, politics, and economics; and is presented in a standardized format (for more information about format, see Defense Language Institute Foreign Language Center, 2009). The test yields separate scores for reading and listening, which are based on the number of questions answered correctly without penalty for incorrect answers. Validation data on a recent version of the test indicate that the test is highly reliable, with individual item reliability of at least .90 and a correlation of at least .85 between the overall test score and the Constructed Response Test, an instrument created specifically for validation of the DLPT (Duterte, 1999, p. 121).

¹ Note that we have anonymized the languages in order to protect the privacy of the language program that provided the data for this study.

Table 1
Questions About Homework on the Student Survey Administered at the End of the Course

Variable	Question	Answer choices
Relevance	How relevant was the assigned homework to what was being taught?	(a) not at all relevant (b) somewhat relevant (c) very relevant (d) cannot determine
Feedback	Please indicate your level of agreement or disagreement with the following statement: "I received useful feedback on homework."	(a) strongly disagree (b) disagree (c) agree (d) strongly agree (e) don't know, no opinion
Grading	Please indicate your level of agreement or disagreement with the following statement: "Homework grading standards were fair."	(a) strongly disagree (b) disagree (c) agree (d) strongly agree (e) don't know, no opinion
Time	On average, how much time did you spend each day outside of class on assigned homework?	(a) none (b) 0.5 hr or less (c) more than 0.5 to 1 hr (d) more than 1 to 1.5 hr (e) more than 1.5 to 2 hr (f) more than 2 to 2.5 hr (g) more than 2.5 hr
Miscellaneous	Do you have any comments about the feedback and grading in the program?	[free response]

The OPI, a test administered by the American Council on the Teaching of Foreign Languages (ACTFL), comprises a "standardized procedure for the global assessment of functional speaking ability" involving an interview in person or over the phone with a trained and certified ACTFL interviewer (American Council on the Teaching of Foreign Languages, 2013). Like the DLPT, the OPI is intended to be a test "independent of any specific curriculum" and, thus, "it is irrelevant when, where, why and under what conditions the candidate acquired his/her speaking ability in the language" (American Council on the Teaching of Foreign Languages, 2013). The interviewer structures the interview to be interactive and adapt to the test candidate's speaking abilities, and he or she discusses a variety of topics. For scoring purposes, the interview is audio recorded and then rated by one or two certified testers according to a standardized set of scoring criteria (described in Breiner-Sanders, Lowe, Miles, & Swender, 2000 and the American Council on the Teaching of Foreign Languages' online guide to the test). Like the DLPT, the OPI is a highly reliable test, with interrater reliability reported to exceed .94 across a variety of languages and a correlation of at least .92 between the official OPI score and judgments of naive (non-ACTFL-certified) native speakers (Henning, 1992, p. 367; Surface & Dierdorff, 2003, p. 512).

Statistical Analyses

To address Question 1, we examined how survey responses about assigned homework were correlated with course outcome measures. The homework variables were HWRelevance, HWFeedback, HWGrading, and HWTime, and the outcome variables were GPA, DLPT-L, DLPT-R, and OPI scores. Both the survey responses and the standardized outcome measures provided ordinal data, which violated the assumptions of normality; therefore, Kendall's tau (τ) correlations were computed instead of Pearson's product-moment (r) correlations.

To address Question 2, we used mixed-effects regression to build hierarchical models of GPA and of the likelihood of attaining test scores indicative of "limited working proficiency" in the FL studied (i.e., DLPT-L, DLPT-R, and OPI scores of at least 2). Hierarchical modeling was performed for these analyses because of the hierarchically nested nature of our data, in which students were nested within classes and classes were nested within languages; furthermore, the treatment of class and language as random, rather than fixed, effects in hierarchical modeling allows for generalization of the findings to a wider population (see, e.g., Patall et al., 2010; Raudenbush & Bryk, 2002). The models were built using the *lmer()* function in R, an open-source statistical package (R Development Core Team, 2013). Because the objective was to examine the effect of homework variables after controlling for background variables, we compared two models of each outcome: Model 1 contained fixed-effect terms for student background variables (e.g., age, sex) and random intercepts for group variables (i.e., class, language), and Model 2 added to Model 1 all of the homework variables (e.g., HWTime, HWRelevance), with random slopes by the grouping variables.

The potential issue of collinearity arose for the three attitudinal homework-related predictors: HWRelevance, HWFeedback, and HWGrading. Although none of these predictors were correlated with HWTime, they were significantly correlated with each other. This was expected, as they were all measuring the same general construct (i.e., attitudes toward homework). However, because we had no theoretical justification for combining these measures into one composite measure (e.g., by averaging them) and were also more concerned with the informativeness of these predictors than with precise coefficient estimates, we kept these predictors separate from each other and left all of them in Model 2 rather than dropping predictors (in line with, e.g., Arceneaux & Huber, 2007). To check whether predictors were likely to have imprecise coef-

ficients due to collinearity, we examined the variance inflation factors.

Results

Psychometric Properties of Survey and Outcome Measures

In order to examine the reliability of the individual survey items addressing attitudes toward homework, we calculated single-item reliability (λ_j) at the class level with Equation 1, where n_j represents the average cluster size (= 14.4 students per class) and ρ_I represents the intraclass correlation between responses in the same cluster (Ginns & Barrie, 2004; Snijders & Bosker, 2012):

$$\lambda = \frac{n_j \rho_I}{1 + (n_j - 1) \rho_I}. \quad (1)$$

The value of the intraclass correlation ρ_I was determined based on variance components derived from a linear regression model (cf. Lessells & Boag, 1987). Each model contained one predictor (Class, as a random effect) and was built with the *lmer()* function in R. The among-groups variance accounted for by the random-effect term, and the remaining within-group variance was used to calculate the value of ρ_I according to Equation 2:

$$\rho_I = \frac{\text{among . groups . variance}}{\text{within . group . variance} + \text{among . groups . variance}}. \quad (2)$$

Using this method, we determined that the HWRelevance, HW-Feedback, and HWGrading measures had acceptable reliability at the class level ($\lambda_j = .75$, $\lambda_j = .78$, and $\lambda_j = .70$, respectively).

Whereas ratings of HWRelevance, HWFeedback, and HWGrading addressed properties of the homework assigned to an entire class (and, therefore, were expected to show consistency within a class), estimates of HWTime were not expected to show consistency within a class, because these had to do with individual students' idiosyncratic homework practices. Consequently, we examined the reliability of the survey item addressing HWTime at the level of the individual student. In order to do so, we analyzed data from a subset of 2,097 participants who responded to this survey item twice: once in the middle of their course and once at the end of their course. Using these data and the same method as before (with a cluster size of 2 responses per student and a linear model with Student as a random effect), we determined that the HWTime measure had acceptable reliability at the individual level ($\lambda_j = .76$).

As for the validity of the survey measures, criterion validity could not be established directly using objective measures of the given variables for two reasons. On the one hand, the general construct measured by the survey items addressing HWRelevance, HWFeedback, and HWGrading (which may be described as "general satisfaction with the homework") was a subjective construct; therefore, it was not possible to measure this construct using objective means for comparison with the survey measures. In the case of HWTime, on the other hand, determining an objective value for time spent would have required surveillance of home-

work completion, and this was not feasible within the context of the language program under study.

Although direct estimates of criterion validity are not available, there is evidence that these measures had solid construct validity. First, the attitudinal measures, which we expected to reflect the construct of Homework Satisfaction, were significantly correlated with each other ($\tau = .31$ to $\tau = .51$, $z > 16.871$, $p < .0001$). These correlations provided evidence of convergent validity, consistent with the hypothesis that these measures were indexing the same general construct (Schacht & Aspelmeier, 2005). Second, the HWTime measure, which we expected to reflect the construct of Time Spent, was not correlated with any of the attitudinal measures ($\tau = -.02$ to $\tau = .02$, $|z| < 1.101$, n.s.), providing evidence of discriminant validity. In other words, the HWRelevance, HW-Feedback, and HWGrading survey items appeared to be measuring aspects of the same construct, whereas the HWTime survey item appeared to be measuring a different construct. Because we have no reason to believe that these survey items were measuring variables other than the ones addressed on the survey, we assume that they measured what they were ostensibly measuring (i.e., satisfaction with the assigned homework and time spent completing it).

With regard to the only survey item addressing an objectively measurable construct (HWTime), it is worth noting that findings from validation studies examining self-report measures of time in other domains suggest that this kind of measure is likely to have acceptable criterion validity at the level required for the purposes of the current study (i.e., the level of *relative* time). For example, studies of sedentary time and physical activity time have repeatedly reported significant and often strong correlations between self-reported values of time and objectively measured values of time (e.g., Clark et al., 2011; Clemes, David, Zhao, Han, & Brown, 2012; Friedenreich et al., 2006; Hayden-Wade, Coleman, Sallis, & Armstrong, 2003; Strath, Bassett, Ham, & Swartz, 2003). Moreover, analyses of time spent using electronic communication have reported strong correlations ($r = .59$ to $r = .87$) between self-reported time spent and actual time spent on Facebook, Twitter, and e-mail (Junco, 2013). What this research shows is that self-reports of time spent tend to be closely related to actual time spent, such that they can provide insight into the relative amount of time spent in comparison to other individuals within a given group.

Three additional facts further suggest that our HWTime measure had acceptable criterion validity for the purposes of the current investigation. Although no study to our knowledge has examined the validity of self-report measures of time spent on homework using objective measures of time spent, the one study that examined the reliability of self-report measures (de Jong et al., 2000) found that, on average, an estimate of time spent obtained via a single-item self-report was less than two times an estimate obtained via multiple entries in a logbook. In addition, at least one validation study of self-report measures of physical activity time has shown that, among youth respondents, self-report measures have higher reliability and validity for older respondents, most likely because of their better memory recall (Sallis, 1991); this age effect suggests that the adult students in our study were probably better able to estimate their time spent via introspective self-report than the adolescents examined in de Jong et al. (2000). Finally, even if we were to assume that our sample of American adults overestimated HWTime by as large a factor as Dutch adolescents

appear to, this would mean that the average error associated with an individual's self-report of HWTime would amount to only one level on the relevant survey item (for choice a, 0; for choice b, 0; for choice c, 0–1; for choice d, 1; for choice e, 1–2; for choice f, 2; for choice g, 0–3). Thus, it is reasonable to believe that the measure of time spent provided by the HWTime survey item had acceptable validity for the present purposes.

In regard to the outcome measures we examined, all showed solid reliability and validity. The reliability of teacher-assigned GPA at the level of the individual student was found to be excellent using the method described above ($\lambda_j = .90$). Furthermore, GPA was positively correlated with all standardized test measures ($\tau = .27$ to $\tau = .46$, $z > 17.331$, $p < .0001$), and the test measures were all positively correlated with each other ($\tau = .24$ to $\tau = .45$, $z > 13.807$, $p < .0001$), providing evidence of convergent validity. These results were consistent with previous descriptions of the DLPT and OPI, which have reported high reliability and validity for these tests. Consequently, we consider students' teacher-assigned grades and DLPT and OPI scores to be suitable for analysis as outcome measures in the present study.

Descriptive Statistics on Homework Variables

Prior to calculating correlations, we examined descriptive statistics on the homework variables to check for systematic differences between languages, especially in relation to difficulty level. Table 2 displays the median, mean, and standard deviation of coded responses to survey items regarding homework for all languages combined, as well as for each language separately. For each variable, a mean and standard deviation are provided in addition to a median to give a sense of the distribution of responses about the median.

The descriptive statistics revealed that responses varied across languages to a limited extent, with median responses not differing by more than one level across languages. Nevertheless, there were a number of significant differences between languages; for example, Language D showed significantly lower HWRelevance than Language B (Wilcoxon $W = 2391.5$, $n_1 = 41$, $n_2 = 271$, $p < .0001$), and Language A showed significantly shorter HWTime than Language C (Wilcoxon $W = 88875.5$, $n_1 = 834$, $n_2 = 315$, $p < .0001$). Notably, however, significant differences between languages did not show any systematic relationship with language difficulty. It was not the case, for instance, that more difficult languages were consistently associated with longer HWTime or higher HWRelevance than less difficult languages. Differences between languages were not related to other factors such as enrollment, alphabet, or grammatical features, either. Thus, we found no evidence of systematic differences between languages in attitudes toward homework or time spent on homework. However, in the analyses that follow we continue to consider by-language results in addition to overall results in order to check the consistency of global patterns in the data.

Correlations Between Aspects of Homework and Course Outcomes

The results of our correlational analyses are shown in Table 3 and Figure 1, which plots the overall relationship between homework variables and the interval outcome variable of GPA. When

Table 2
Descriptive Statistics on Coded Homework Variables

Language	<i>n</i>	Variable	Median	<i>M</i> (<i>SD</i>)
A–G	2,342	HWRelevance	1	1.33 (0.59)
		HWFeedback	2	2.05 (0.80)
		HWGrading	2	2.26 (0.70)
		HWTime	3	3.47 (1.47)
A	834	HWRelevance	1	1.23 (0.56)
		HWFeedback	2	1.93 (0.80)
		HWGrading	2	2.17 (0.71)
		HWTime	3	3.10 (1.40)
B	271	HWRelevance	2	1.59 (0.56)
		HWFeedback	2	2.32 (0.65)
		HWGrading	3	2.42 (0.66)
		HWTime	4	3.92 (1.33)
C	315	HWRelevance	1	1.44 (0.56)
		HWFeedback	2	2.34 (0.70)
		HWGrading	2	2.35 (0.68)
		HWTime	4	3.99 (1.58)
D	41	HWRelevance	1	0.90 (0.54)
		HWFeedback	1	1.15 (0.99)
		HWGrading	2	1.44 (0.90)
		HWTime	4	3.83 (1.32)
E	324	HWRelevance	1	1.12 (0.59)
		HWFeedback	2	1.81 (0.80)
		HWGrading	2	2.08 (0.71)
		HWTime	3	3.42 (1.41)
F	406	HWRelevance	1	1.45 (0.56)
		HWFeedback	2	2.14 (0.76)
		HWGrading	2	2.41 (0.58)
		HWTime	4	3.60 (1.44)
G	151	HWRelevance	1	1.46 (0.56)
		HWFeedback	2	2.21 (0.76)
		HWGrading	3	2.46 (0.66)
		HWTime	3	3.35 (1.48)

Note. Statistics were derived from responses to course-end survey items. Range for HWRelevance = 0–2; for HWFeedback and HWGrading = 0–3; for HWTime = 0–6. HW = assigned homework.

all languages were combined, every homework variable correlated significantly with every outcome variable at $p < .001$ (after application of the Bonferroni correction for multiple comparisons), with one exception: HWTime in relation to OPI scores. This latter correlation, however, was negative like the other correlations of HWTime with outcome variables. To check the consistency of the overall patterns across different languages, correlations were also calculated by language, which revealed that the relationship of homework variables to outcome variables was highly consistent across languages. Although correlations within each language were not all statistically significant, they mostly showed the same direction of effect as the overall correlations, and within-language correlations that differed in direction from the overall correlations were all nonsignificant.

As predicted, HWRelevance, HWFeedback, and HWGrading were positively correlated with outcomes. This was the case for all outcome measures; across all languages; and within each language, with few exceptions. Out of the 84 within-language correlations of attitudinal variables and outcome variables, there were only seven exceptions to the overwhelming pattern of positive correlation.

In contrast, HWTime was negatively correlated with outcomes. This, too, was the case for all outcome measures, both overall and within each language. Notably, there were no exceptions to this

Table 3
Nonparametric Correlations Between Homework and Outcome Variables

Language (n)	Variable	GPA	DLPT-L	DLPT-R	OPI
A–G (2,342)	HWRelevance	.08***	.10***	.12***	.15***
	HWFeedback	.07***	.12***	.10***	.14***
	HWGrading	.09***	.10***	.10***	.13***
	HWTime	-.13***	-.16***	-.13***	-.02
A (834)	HWRelevance	.06*	.08*	.07*	.13***
	HWFeedback	.03	.07*	.03	.09**
	HWGrading	.08**	.06	.07*	.12***
	HWTime	-.18***	-.18***	-.14***	-.04
B (271)	HWRelevance	.08	.13*	.07	.15*
	HWFeedback	.06	.12*	.05	.18**
	HWGrading	.05	.05	-.02	.06
	HWTime	-.12*	-.16**	-.14**	-.09
C (315)	HWRelevance	.06	.20***	.16**	.10
	HWFeedback	.09*	.19***	.15***	.08
	HWGrading	.14**	.14**	.09	.15**
	HWTime	-.11**	-.15**	-.07	-.11*
D (41)	HWRelevance	-.01	.17	.35*	.15
	HWFeedback	.03	.29*	.25	-.01
	HWGrading	.12	.22	.36**	.23
	HWTime	-.10	.15	-.22	-.17
E (324)	HWRelevance	.12**	.13**	.17***	.01
	HWFeedback	.09*	.15***	.18***	.05
	HWGrading	.12**	.11*	.17***	.12*
	HWTime	-.17***	-.17***	-.19***	-.04
F (406)	HWRelevance	.06	.00	.04	.03
	HWFeedback	.10**	.09*	.04	.08
	HWGrading	.09*	.07	.04	.03
	HWTime	-.15***	-.19***	-.16***	-.05
G (151)	HWRelevance	.07	.01	.05	-.05
	HWFeedback	-.01	.11	.00	-.10
	HWGrading	.00	.09	.04	-.06
	HWTime	-.15*	-.21**	-.15*	-.22**

Note. Statistics were derived from responses to course-end survey items. Range for HWRelevance = 0–2; for HWFeedback and HWGrading = 0–3; for HWTime = 0–6. HW = assigned homework; GPA = grade point average; DLPT = Defense Language Proficiency Test, with subtests L (listening) and R (reading); OPI = Oral Proficiency Interview.
* $p < .05$. ** $p < .01$. *** $p < .001$ (after Bonferroni correction).

pattern; all 28 within-language correlations of HWTime and outcome variables were negative. Furthermore, the negative correlations were consistent across different parts of the HWTime scale. That is to say, HWTime was negatively correlated with outcomes both at relatively low levels of HWTime (Levels 0–3: 0–1.5 hr/day) and at relatively high levels of HWTime (Levels 4–6: >1.5 hr/day). For example, HWTime showed a significant negative correlation with final GPA both for HWTime \leq 1.5 hr/day ($\tau = -.11, z = -5.062, p < .0001$) and for HWTime > 1.5 hr/day ($\tau = -.05, z = -2.200, p < .05$). This was also the case with DLPT-R scores (HWTime \leq 1.5 hr/day: $\tau = -.09, z = -3.695, p < .001$; HWTime > 1.5 hr/day: $\tau = -.06, z = -2.518, p < .05$). These results suggested that the observed negative correlation between HWTime and course outcomes was a robust relationship not specific to particular amounts of time spent on homework.

In short, our correlational analyses demonstrated that better ratings of HWRelevance, HWFeedback, and HWGrading were associated with more favorable course outcomes, whereas increased HWTime was associated with less favorable course outcomes. These analyses, however, abstracted away from the effects

of possible intermediate variables such as language aptitude. For this reason, we also conducted hierarchical regression analyses, which are discussed in the next section.

Regression Models of Course Outcomes

As described above, we took a hypothesis-testing approach to our regression analyses, comparing two models for each outcome variable. Model 1 served as a base model containing all predictors except the homework variables; Model 2 added in the homework variables. The full set of predictors comprised two group variables, five student background variables, and four homework-related variables. Group variables were FL studied (Language: A–G) and specific class (Class: 163 values). Background variables were language learning aptitude as measured by the DLAB (Aptitude: 85–151), age upon completion of the course-end survey (Age: 18–50), sex (Sex: male or female, base level female), level of education (Education: 1–9), and level of motivation for learning the FL studied (Motivation: 0–4). Homework variables were HWRelevance (0–2), HWFeedback (0–3), HWGrading (0–3), and HWTime (0–6). At the suggestion of an anonymous reviewer, the homework variables, along with Education and Motivation, were treated as continuous predictors in our models so as to produce one overall coefficient per predictor in the model output.

The two outcome variables were final GPA and the likelihood of attaining limited working proficiency (LWP) in all three modalities tested. For both outcomes, collinearity between predictor variables was not a problem in the final model, as indicated by variance inflation factors that were all well below 2. Furthermore, the errors were found to be independent in a Durbin–Watson test and to have a normal distribution, as shown in a histogram and a Q–Q plot (Field, Miles, & Field, 2012). The residuals were only slightly heteroscedastic and did not require correction (cf. Fox, 1997). In addition, we confirmed that the continuous predictors in each model were linearly related to the logit of the outcome variable by examining the contribution to the model of the interaction between the continuous predictor and its log transformation (Hosmer & Lemeshow, 2000). In exploratory model building, we found no significant interactions between predictors for either outcome; consequently, we included only simple effects in the comparisons of Model 1 and Model 2.

To determine the appropriate level at which to consider the homework variables in modeling, we investigated how much of the variance in each of the homework variables was attributable to student versus class versus language. The relative amount of variance accounted for by these factors was examined by comparing three single-predictor models of a given homework variable, which differed only in whether Student, Class, or Language was entered as a random effect; in all cases, the outcome (homework variable) data comprised midcourse and course-end survey responses from the subset of participants who had completed both surveys. The results of this analysis showed that each of the homework variables was best analyzed as a student-level variable, as Student consistently accounted for more variance than either Class or Language. Consequently, all models had student background variables and homework variables entered at Level 1, with random intercepts for group variables (i.e., class, language) entered at Level 2 and Level 3.

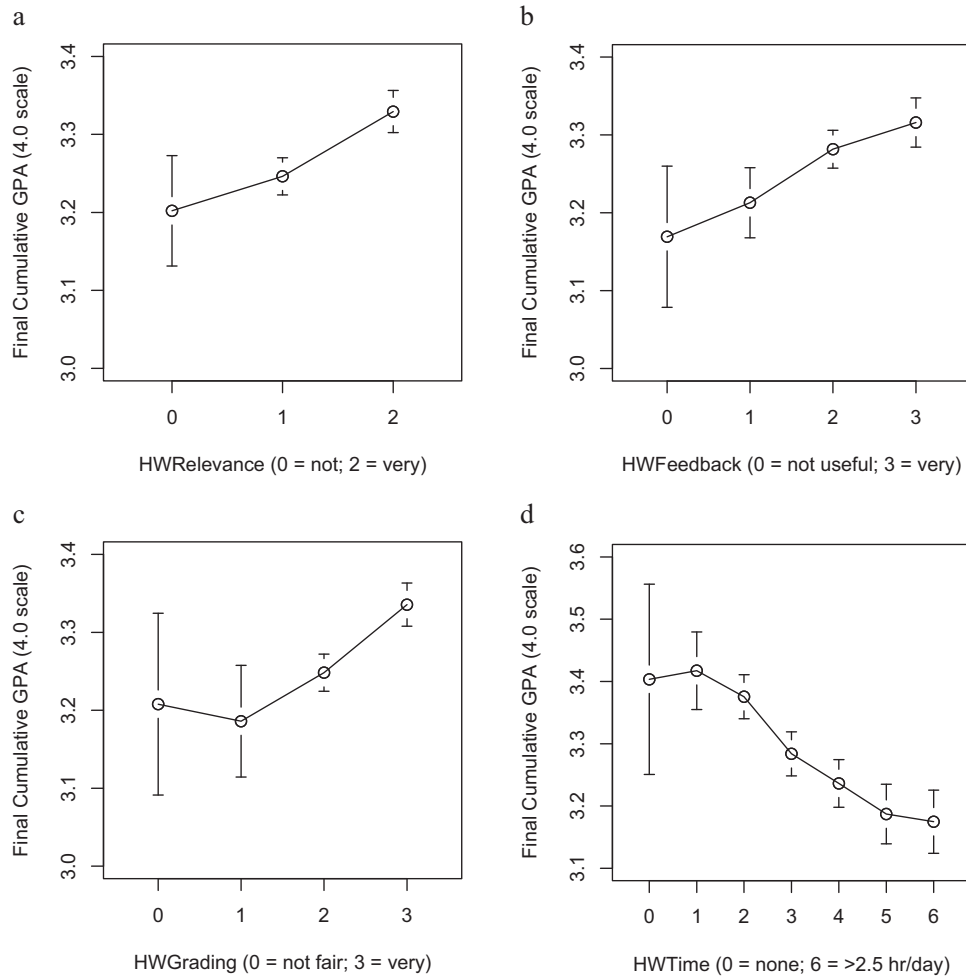


Figure 1. Variation in final cumulative GPA by (a) HWRelevance, (b) HWFeedback, (c) HWGrading, and (d) HWTime. Each point and error bar plot, respectively, the mean and 95% confidence interval over participants. GPA = grade point average; HW = assigned homework.

For the outcome of final GPA, Model 2 significantly improved upon the predictions of Model 1, $\chi^2(12) = 144.29$, $p < .0001$, lowering the Akaike information criterion (AIC, a measure of the information lost when a model is used to describe a set of data; Akaike, 1974) from 2,372.4 to 2,252.1. The output of GPA Model 2 is summarized in Table 4. As expected, Aptitude, Education, and Motivation were positively related to GPA. However, there was no significant effect of Age or Sex. As for the homework-related variables, although there was no significant effect of HWFeedback, HWGrading and HWRelevance were both positively related to GPA. HWTime, on the other hand, was negatively related to GPA. Post hoc comparisons of GPA distributions by level of HWTime revealed that a significant difference in GPA between students who reported spending some amount of time on assigned homework and those who reported spending no time on assigned homework emerged at values of HWTime of more than 1.5 hr/day ($ts > 2.16$, $ps < .05$).

In the case of LWP, too, Model 2 significantly improved upon the predictions of Model 1, $\chi^2(12) = 51.07$, $p < .0001$, lowering the AIC from 2,802.6 to 2,775.5. The output of LWP Model 2 is

summarized in Table 5. The effects of the predictor variables here were similar overall to those in the final model of GPA. As in the model of GPA, the coefficients for Aptitude and Education were both significantly greater than zero, whereas that for Sex was not significantly different from zero. The coefficient for Motivation, although positive, was also not significantly different from zero. In contrast, the coefficient for Age was negative (unlike in the model of GPA). Coefficients for all homework-related variables were significantly different from zero: Those for HWGrading, HWFeedback, and HWRelevance were positive, whereas that for HWTime was negative.

Inspection of the random effects in GPA Model 2 and LWP Model 2 revealed that there was overall little variance in the effects of homework-related variables across levels of the grouping variables (see Tables 4–5). For both models, the greatest variance was in the effects of HWFeedback and HWRelevance across different classes. There was also variance in the effect of HWTime across different languages but not enough to change the polarity of the coefficient, which was negative across all languages. All other variances were near zero.

Table 4
Final Linear Model (Model 2) of GPA

Predictor	Fixed effect			Random effect			
	β	SE	t	Groups	Predictor	Variance	SD
(Intercept)	3.175	0.052	60.75***	Class	(Intercept)	0.0016	0.0404
Aptitude	0.009	0.001	11.91***	Class	HWRelevance	0.0017	0.0416
Education	0.022	0.005	4.22***	Class	HWFeedback	0.0006	0.0238
Motivation	0.028	0.008	3.26**	Class	HWGrading	0.0000	0.0000
Age	0.004	0.002	1.49	Class	HWTime	0.0002	0.0132
Sex: male	-0.021	0.019	-1.11	Language	(Intercept)	0.0000	0.0000
HWRelevance	0.036	0.016	2.23*	Language	HWRelevance	0.0000	0.0000
HWFeedback	0.013	0.013	1.00	Language	HWFeedback	0.0000	0.0047
HWGrading	0.033	0.014	2.42*	Language	HWGrading	0.0000	0.0000
HWTime	-0.055	0.014	-3.98***	Language	HWTime	0.0011	0.0325
				Residual		0.1425	0.3775

Note. SE = standard error; SD = standard deviation; HW = assigned homework; GPA = grade point average.
* $p < .05$. ** $p < .01$. *** $p < .001$.

Discussion

This article reported the results of the first systematic investigation of relationships between properties of assigned homework and FL course outcomes. We found that students' perceptions of the relevance of homework, the usefulness of homework feedback, and the fairness of homework grading were positively correlated with outcomes, whereas reported time spent on homework was negatively correlated with outcomes. The findings of our regression analyses further showed that homework variables continued to display these relationships with outcomes after the effects of several background variables were taken into account, suggesting that the observed correlations between homework variables and outcome variables were the product of legitimate links between homework variables and outcomes, rather than the artifact of relationships with a covariate.

Before we discuss these results further, some limitations of this study should be noted. The homework variables were based on self-reported data from students, which may have suffered from recall bias (cf. de Jong et al., 2000). Our data, moreover, represented only a subset of seven languages over a select time period.

Furthermore, although we controlled for several background variables in our models, we did not have other information relevant to course outcomes, such as amount of time spent on nonassigned (i.e., self-regulated) study, nor were we able to gather this information, as the data used were gathered from standard surveys that we were not in a position to change. Despite these limitations, however, the current findings provide compelling evidence of relationships between homework variables and FL course outcomes.

Below we present an assessment of our findings and what we regard as the most plausible interpretation. We first evaluate the results in light of previous results and then discuss explanations for the two major patterns: the positive relationship between attitudes and outcomes and the negative relationship between time spent and outcomes. Although we consider a number of possible accounts, it is important to keep in mind that our discussion is not exhaustive. That is to say, we give space here only to those accounts we regard as most likely given the literature on homework, the literature on language learning, and our knowledge of the language program under study.

Table 5
Final Logistic Model (Model 2) of Attaining Limited Working Proficiency (LWP) in Three Modalities (i.e., DLPT-L, DLPT-R, and OPI Scores of at Least 2)

Predictor	Fixed effect			Random effect			
	β	SE	z	Groups	Predictor	Variance	SD
(Intercept)	-1.988	0.462	-4.300***	Class	(Intercept)	0.0763	0.2763
Aptitude	0.029	0.005	6.332***	Class	HWRelevance	0.0452	0.2126
Education	0.098	0.032	3.106**	Class	HWFeedback	0.0009	0.0298
Motivation	0.053	0.051	1.047	Class	HWGrading	0.0000	0.0000
Age	-0.070	0.016	-4.511***	Class	HWTime	0.0000	0.0000
Sex: male	0.202	0.113	1.788	Language	(Intercept)	0.7465	0.8640
HWRelevance	0.249	0.094	2.638**	Language	HWRelevance	0.0000	0.0002
HWFeedback	0.163	0.077	2.111*	Language	HWFeedback	0.0000	0.0000
HWGrading	0.184	0.084	2.181*	Language	HWGrading	0.0000	0.0000
HWTime	-0.118	0.037	-3.232**	Language	HWTime	0.0005	0.0233

Note. SE = standard error; SD = standard deviation; HW = assigned homework; DLPT = Defense Language Proficiency Test, with subtests L (listening) and R (reading); OPI = Oral Proficiency Interview.
* $p < .05$. ** $p < .01$. *** $p < .001$.

Assessment of the Findings

Comparisons of the present results with prior results suggest that the relationship between homework variables and outcomes in the present study was as strong as, if not stronger than, the relationship found in prior work. For example, whereas Cooper et al. (1998) reported parametric correlations between attitudinal variables and outcome variables as large as $|r| = .11$, our nonparametric correlations between attitudinal variables and outcome variables ranged up to $|\tau| = .35$. In fact, 33 of 96 (34%) were larger in magnitude than $|\tau| = .11$. Similarly, whereas Cooper et al. (1998) reported correlations between time spent and outcome variables as large as $|r| = .17$, our correlations between time spent and outcome variables ranged up to $|\tau| = .22$, and 7 of 32 (22%) were larger in magnitude than $|\tau| = .17$. Because converting our Kendall's correlations to Pearson's correlation equivalents would only increase the magnitude of our results (see Walker, 2003 on Kendall's-to-Pearson's conversion), this is strong evidence that our data evince a relatively robust relationship between homework variables and outcomes.

In addition, our regression models compared favorably to previous models of FL proficiency measures. Although calculation of a traditional r^2 is not straightforward for models with random effects, using the method devised by Nakagawa and Schielzeth (2013) we estimated the marginal r^2 and conditional r^2 of the final GPA model to be, respectively, .16 and .19, which are both higher than the average adjusted r^2 of .14 reported in Lett and O'Mara (1990). Moreover, with just fixed effects the model was able, on average, to predict a student's GPA within about 0.3 of the actual GPA. The final LWP model was also relatively informative, with Hosmer-Lemeshow, Cox-Snell, and Nagelkerke pseudo r^2 values (Field et al., 2012) of .21, .24, and .32, respectively, and a rate of correct predictions of approximately 72%. These facts suggest that the final regression models accounted relatively well for variation in outcomes.

Of note, the effects of homework variables not only were statistically significant but also were sizable relative to the effects of background variables. In exploratory modeling, we examined the relative informativeness of each individual fixed-effect predictor by building separate regression models containing single predictors and then comparing the AIC associated with these different models. These comparisons revealed that the homework variables were often among the most informative predictors of outcome variables. Out of nine total fixed-effect predictors, HWTime, for example, was the second-most informative predictor of GPA and of attaining LWP, and HWGrading was the third-most informative predictor of attaining LWP and the fourth-most informative predictor of GPA. Thus, our final models, together with the results of our exploratory modeling, provide strong evidence of robust relationships between homework variables and FL course outcomes. Next, we turn to the interpretation of these relationships, beginning with the positive relationship of attitudinal variables with outcomes.

The Positive Relationship Between Attitudes and Outcomes

The positive relationships of HWRelevance, HWFeedback, and HWGrading with outcomes show that, consistent with the findings

of Cooper et al. (1998), attitudes toward homework can relate to outcomes (in this case, with adult learners of an FL). If it can be assumed that students' attitudes toward homework are formed at least somewhat independently from their performance in a course, our results suggest that positive attitudes toward homework have a positive effect on FL course outcomes. However, it is wise to consider whether this assumption is valid, because it is also possible that students' attitudes toward homework are influenced by their course performance: Students who are doing well in a course may be more likely to view the homework positively (under the reasoning that if they are doing well, the homework must be working), whereas students who are doing poorly may be more likely to view the homework negatively (attributing their poor performance to deficiencies in the homework). Such dependence of attitudes on course performance (i.e., causality in the other direction) could also result in the observed relationship between attitudes and outcomes.

Although we cannot say for sure whether or not students' ratings of homework in our study were systematically biased by their own course performance, several facts suggest it might be reasonable to assume that their ratings were, to some degree, independent of course performance. For one, the survey items about homework to which students responded were not general questions about whether the homework was "good" or "bad" but targeted questions about specific aspects of the homework (namely, relevance, feedback, and grading), which might be more likely to elicit impartial judgments not based on course performance than questions asking for a general evaluation of the homework. In addition, a closer look at the data suggests that the positive correlations between ratings and outcomes were not simply due to rating disparities between students doing well and students doing poorly. For example, when the data set is split into halves by GPA (below or above average), it is apparent that in both halves the entire rating scale is used with a similar distribution and that positive correlations of HWRelevance, HWFeedback, and HWGrading with GPA persist (reaching significance in four of six cases and approaching significance in the other two).

Student comments on the course-end survey provide further support for the assumption that ratings of homework were not determined by course performance. Approximately 20% of commenters expressed dissatisfaction with the consistency and transparency of grading, although many of these students were in fact receiving high grades. For example, one student expressed skepticism about the way his work was graded, even though his grades were actually good:

I found that the quality of my homework did not affect the grade I received on assignments. Because I received high marks on homework at the beginning of the course, I received them for the whole course even if I did not deserve them on certain assignments.

Another student commented on the inconsistency in grading for other students, while he himself received high grades:

I had an excellent counselor who graded my homework and often gave me feedback. However, I would have friends who would receive lesser grades while having done the same amount of work that I did. It's not fair to those students to receive a "B" when I received an "A." There needs to be more uniformity in grading the homework.

These kinds of comment provide evidence that students were able to respond impartially to the survey items; in particular, they appeared able to evaluate the homework independent of their own course performance.

Together, the specificity of the survey items about homework, the similarity in rating distribution between high-performing students and low-performing students, and students' qualitative descriptions of the homework suggest that it is probably safe to assume that the observed student attitudes toward assigned homework were not caused by their interim outcomes. Given this assumption, our findings could potentially be interpreted as evidence that positive attitudes toward homework tend to boost outcomes, but we believe that the link between attitudes and outcomes is probably not so direct. The mechanism behind this link is not something the current study was designed to examine; however, it is important to emphasize that our measure of attitudes was based not on students' general feelings about homework but rather on specific qualities of the homework that are likely to be beneficial for FL learning: relevance, usefulness of feedback, and fairness in grading. Some studies have suggested, for example, that the benefits of FL homework are modulated by the quality of feedback provided (Cardelle & Corno, 1981; Wallinger, 2000). Thus, to be precise, our findings should be interpreted as evidence that certain subjective qualities of homework are positively linked to FL course outcomes. These qualities of homework are likely to influence both attitudes and performance, which may constitute a mechanism behind previously documented links between attitudes and outcomes. In fact, this would be quite reasonable, because attitudes are not usually random but are based on something. To put it another way, positive attitudes toward homework may be linked to better outcomes at least in part because they are indexical of positive qualities of the homework that support higher achievement.

The Negative Relationship Between Time Spent and Outcomes

Whereas the positive link between outcomes and HWRelevance, HWFeedback, and HWGrading is unsurprising, the negative link between outcomes and HWTime contradicts many of the findings in other subject areas and merits further comment. Although it is not clear why spending more time on assigned homework is associated with poorer outcomes, here we discuss three possible explanations. These accounts differ in terms of the proposed mechanism behind the covariation of HWTime and outcome variables; however, they are similar in that they are all noncausal. We do not consider it likely that doing assigned homework directly causes students' performance in an FL course to suffer; positing such an effect would go against the logic that "practice makes perfect" and call into question the very purpose of doing homework. Instead, we have identified the three most probable accounts for this counterintuitive result: a progress-to-time account, an ability-to-time account, and an opportunity cost account.

The first account of the negative relationship between HWTime and outcomes posits that the negative link arises due to disparities in HWTime following from disparities in students' progress in a course. That is, students who are doing well in a course may spend less time on homework because they feel little need to do so given the positive feedback they receive on their progress, whereas

students who are doing poorly in a course may spend more time on homework because they themselves feel the need to do so or are otherwise advised to do so in order to improve. Thus, students' interim progress leads to adjustments in HWTime, considered here as the progress-to-time account. This account is noncausal because in this scenario HWTime does not cause outcomes to get better or worse. On the contrary, HWTime fails to influence outcomes one way or the other. That is to say, students who are doing well are ultimately not hurt by spending relatively little time on homework; conversely, students who are doing poorly are ultimately not helped by spending a lot of time on homework. After all, homework is just one part of an intensive FL course, and there are many other variables that are likely to contribute to students' success.

The second account of the negative relationship between HWTime and outcomes attributes the negative link to disparities in HWTime following from disparities in students' underlying linguistic ability: High-ability students who are predisposed to master a given FL with ease tend to require less time completing assigned homework than low-ability students who are predisposed to struggle with the FL. Thus, differences in students' language learning ability lead to differences in HWTime, considered here as the ability-to-time account. Under this account the relationship between HWTime and outcomes is epiphenomenal, an artifact of a negative relationship between HWTime and linguistic ability. Ability is, therefore, the main predictor here, accounting for variation both in HWTime and in outcomes. This reasoning is similar to the logic behind the findings of de Jong et al. (2000), which suggested that a negative relationship between math achievement and time spent on math homework was based on prior math knowledge, not time spent per se.

Finally, the third, opportunity cost account of the negative relationship between HWTime and outcomes posits that the amount of time spent on assigned homework can affect outcomes in a counterintuitive manner by way of preventing students from spending time on activities that would help them to perform better on the given outcome measures. In other words, spending time on homework may hinder, rather than help, students, because assigned homework imposes an opportunity cost of not engaging in tailored and/or self-regulated study activities that may provide a greater benefit to students' performance on these measures. This does not exclude the possibility that assigned homework may contribute positively toward other FL learning goals that are not captured in the given outcome measures. However, the basic argument in this account is that traditional homework (i.e., homework assigned generically to a class at large) is not particularly helpful with respect to these outcome measures, and, therefore, spending a lot of time on this kind of homework instead of more beneficial study activities is detrimental to outcomes.

Evidence that may help to clarify the role of HWTime in outcomes can be gleaned from some additional analyses of the data. The first analysis shows that change in reported HWTime from midcourse surveys to course-end surveys is not significantly correlated with change in GPA between any two sections of the course [$|\tau| < .04$, $|z| < 1.87$, n.s.]. Under the progress-to-time account, we would expect changes in students' interim GPA, as an indication of their progress in a course, to result in systematic changes to the amount of time they spend on homework; in particular, we would expect students to devote more time to homework when they see that they are doing poorly and/or less

time to homework when they see that they are doing well. The fact that these correlations are not found is, therefore, inconsistent with the progress-to-time account. The second analysis shows that HWTime is still a significant predictor of one or more outcome variables when only students with similar levels of language aptitude (i.e., similar DLAB scores) are examined. For example, when only students with exceptionally high DLAB scores of 140 or above (the top 5%) are considered, HWTime is still negatively correlated with final GPA and DLPT-L scores ($\tau < -.18$, $z < -2.153$, $p < .05$); when only students with the lowest DLAB scores (the bottom 4%, with DLAB scores of less than 100) are considered, HWTime is still negatively correlated with DLPT-L scores ($\tau = -.29$, $z = -3.262$, $p < .01$); and when only students with average DLAB scores (the middle 10%, with DLAB scores of 115–117) are considered, HWTime is still negatively correlated with final GPA, DLPT-L, and DLPT-R scores ($\tau < -.11$, $z < -2.197$, $p < .05$). Note that these correlations are not simply due to students of different aptitude levels differing in terms of their minimum HWTime, because survey responses from all the above subgroups span the entire range. Under the ability-to-time account, these correlations are unexpected, because variation in HWTime is supposed to be linked to outcomes through variation in aptitude. Consequently, when aptitude is controlled, correlations between HWTime and outcome variables should disappear. The fact that such correlations instead persist is, therefore, inconsistent with the ability-to-time account.

However, the ability-to-time account could still be true if there were another component of “ability” that has not been accounted for. That is, perhaps the correlations between HWTime and outcome variables persist within narrow bands of language aptitude because there is an additional variable—something besides language aptitude and the other background factors included in our models—that affects both HWTime and outcomes and shows significant variation across students with the same language aptitude. The most likely candidate for such a variable would be general (nonlinguistic) cognitive ability. However, there are two reasons to doubt that variation in general cognitive ability is behind the observed relationship between HWTime and outcomes. First, domain-general cognitive abilities such as logical reasoning and working memory do not vary orthogonally to language aptitude; on the contrary, they are widely believed to be central components of language aptitude (for more on this, see [Linck et al., 2013](#) and references cited therein). Consequently, a set of individuals identified as having the same language aptitude is, a priori, biased to show relatively limited variation in general cognitive ability. Second, to be able to enroll in the given language program, students had to pass the equivalent of an IQ test. Although we did not have access to these test scores, this type of requirement would have also constrained the range of general cognitive ability in our study sample. Note that these arguments apply as well to other constructs that could conceivably affect both HWTime and outcome measures (e.g., “test-taking ability”). It is difficult to imagine such constructs would not also be tapped by a language aptitude test like the DLAB.

Thus, additional analyses of the data are most consistent with the opportunity cost account of the negative relationship between HWTime and outcomes. Recall that this account is based on an inverse relationship between time spent on assigned homework and time spent on other, self-regulated study activities. Because

students were not asked about self-regulated study, it is not possible to show conclusively that those who spent less time on assigned homework spent more time on self-regulated study. On the other hand, survey comments suggest that students who spent more time on assigned homework spent less time on self-regulated study and, more generally, that the assigned homework was not especially helpful. The 14% of commenters who expressed opinions of the assigned homework that were at least partly positive were far outnumbered by the 88% who expressed negative opinions, which went well beyond the issues of feedback and grading that were the subject of the free-response question. In particular, 8% of commenters observed that the assigned homework took away from time available for other language exposure (e.g., more authentic input) and/or self-regulated study, and 7% stated that their assigned homework just seemed to be busy work. In addition, 3% expressed the opinion that the assigned homework was not useful because it was not individually tailored, such that 6% questioned the pedagogical value of the assigned homework outright. Several of the most common themes are exemplified in the following student comment:

With the current homework load, students are often overwhelmed with various exercises from the homework book, various packets and handouts, and different listening exercises. There is simply too much to be able to concentrate on doing any one thing well, and many students end up rushing through most of the exercises to be able to complete them all on time. It ends up defeating the purpose of the homework because nothing is well understood or done well, and the student hasn't spent any time just studying vocabulary or reading grammar notes. If the homework was cut in half, the student wouldn't necessarily only spend half the time on it—they would instead not feel such pressure, and they would be able to really think through the exercises that help them to learn. And there would be time to study.

This type of sentiment underscores the potential usefulness of evaluating the relative benefits of structuring out-of-class time with assigned homework activities versus allowing students freedom to pursue activities they identify as helpful for their own learning. Ultimately, without more information on how students use their out-of-class time we can only speculate as to how those who spend little time on assigned homework manage to do well. However, students' own words provide a reasonable explanation for how those who spend a lot of time on assigned homework end up doing poorly: Assigned homework can be counterproductive insofar as it prevents students from making effective use of their out-of-class time.

To be clear, we do not mean to imply that assigned homework does not have any benefits or that FL teachers should stop assigning homework. Our findings simply show that “more” is not necessarily “better”; in fact, it can be worse, as expressed in the above student comment. Although we remain agnostic on the benefits of assigned homework for FL course outcomes, we believe it is important in the structuring of study time outside the classroom to consider the findings of [Patall and colleagues \(Patall et al., 2008, 2010\)](#), which show that offering students choices in homework is associated with more favorable outcomes. As discussed at the beginning of this article, two reasons for this positive effect of choice are the increased feelings of autonomy and intrinsic motivation associated with choice. These factors—in addition to the ability of adult learners to identify activities that are helpful

specifically for their own learning—argue in favor of assigning homework in such a way as to allow students more individualized use of out-of-class time (e.g., tailored exercises, choices in exercises), including time for self-regulated study.

Conclusion

The findings reported in this article are relevant for FL course administration and also point out avenues of future research into FL homework. Given the positive relationship of perceived relevance of homework, usefulness of feedback, and fairness of grading to course outcomes, developing more rigorous course standards that improve student attitudes toward homework may result in improved outcomes as well as overall student satisfaction. For example, implementing a rubric for providing feedback and grading homework that is consistent across teachers could help manage student expectations and thus improve students' perceptions regarding usefulness and fairness. In addition, collecting information about the amount of time students spend on self-regulated study (as opposed to assigned homework) and the types of self-regulated study activities preferred by students would contribute to our knowledge of what out-of-class learning activities students find effective. Along these same lines, implementing a simple intervention study in which students complete generically assigned homework, spend an equivalent amount of time on tailored homework, or use this time on their preferred method of self-regulated study would allow for a comparison of the effectiveness of these kinds of out-of-class learning. This type of insight would greatly improve our understanding of the efficacy of assigned homework and best practices surrounding the use of out-of-class time.

In closing, we would like to point out that although our findings should be interpreted with caution (as our investigation was a correlational study, not an intervention study), they speak to the need for language learners to be empowered to address their own idiosyncratic language deficiencies. Outside of class, some learners may benefit from additional grammar drills, for example, whereas others may already have a solid command of grammar and benefit more from additional speaking practice. Consequently, there may be limited value in a one-size-fits-all approach to out-of-class learning that requires all learners to train all language components. Because it may not be practical for instructors to design an individualized study program with tailored homework for each student, a productive approach may be to assign less required homework and offer suggestions for what students may find useful to do during their remaining time. Although adolescents may still need a teacher to tell them how best to improve their language capabilities, adult learners may be more aware of their language needs and, thus, better prepared to benefit from greater autonomy in using their limited time outside of the classroom to complement their in-class learning.

References

- Akaike, H. (1974). A new look at the statistical model identification. *IEEE Transactions on Automatic Control*, *19*, 716–723. doi:10.1109/TAC.1974.1100705
- American Council on the Teaching of Foreign Languages. (2013). *Testing for proficiency*. Retrieved from <https://www.actfl.org/professional-development/certified-proficiency-testing-program/testing-proficiency>
- Arceneaux, K., & Huber, G. A. (2007). What to do (and not do) with multicollinearity in state politics research. *State Politics and Policy Quarterly*, *7*, 81–101. doi:10.1177/153244000700700105
- Berdan, S. (2012). The world has changed. *The New York Times*. Retrieved from <http://www.nytimes.com>
- Berninger, V. W., Nielsen, K. H., Abbott, R. D., Wijsman, E., & Raskind, W. (2008). Gender differences in severity of writing and reading disabilities. *Journal of School Psychology*, *46*, 151–172. doi:10.1016/j.jsp.2007.02.007
- Brantmeier, C., Schueller, J., Wilde, J. A., & Kinginger, C. (2007). Gender equity in foreign and second language learning. In S. S. Klein (Ed.), *Handbook for achieving gender equity through education* (2nd ed., pp. 305–334). Mahwah, NJ: Erlbaum.
- Breiner-Sanders, K. E., Lowe, P., Jr., Miles, J., & Swender, E. (2000). ACTFL proficiency guidelines—speaking. *Foreign Language Annals*, *33*, 13–18. doi:10.1111/j.1944-9720.2000.tb00885.x
- Burling, R. (1981). Social constraints on adult language learning. In H. Winitz (Ed.), *Annals of the New York Academy of Sciences: Vol. 379. Native language and foreign language acquisition* (pp. 279–290). doi:10.1111/j.1749-6632.1981.tb42015.x
- Byrne, J. A. (2013). Speaking 2 or more languages becoming essential skill. *The New York Post*. Retrieved from <http://nypost.com/>
- Cardelle, M., & Corno, L. (1981). Effects on second language learning of variations in written feedback on homework assignments. *TESOL Quarterly*, *15*, 251–261. doi:10.2307/3586751
- Clark, B. K., Thorp, A. A., Winkler, E. A. H., Gardiner, P. A., Healy, G. N., Owen, N., & Dunstan, D. W. (2011). Validity of self-reported measures of workplace sitting time and breaks in sitting time. *Medicine & Science in Sports and Exercise*, *43*, 1907–1912.
- Clark, J. L. D., & Clifford, R. T. (1988). The FSI/ILR/ACTFL proficiency scales and testing techniques: Development, current status, and needed research. *Studies in Second Language Acquisition*, *10*, 129–147. doi:10.1017/S0272263100007270
- Clemes, S. A., David, B. M., Zhao, Y., Han, X., & Brown, W. (2012). Validity of two self-report measures of sitting time. *Journal of Physical Activity & Health*, *9*, 533–539.
- Cooper, H., Lindsay, J. J., Nye, B., & Greathouse, S. (1998). Relationships among attitudes about homework, amount of homework assigned and completed, and student achievement. *Journal of Educational Psychology*, *90*, 70–83. doi:10.1037/0022-0663.90.1.70
- Cooper, H., Robinson, J. C., & Patall, E. A. (2006). Does homework improve academic achievement? A synthesis of research, 1987–2003. *Review of Educational Research*, *76*, 1–62. doi:10.3102/00346543076001001
- Davidson, J. (2012). Government has foreign language deficit. *The Washington Post*. Retrieved from www.washingtonpost.com/
- Deci, E. L., & Ryan, R. M. (1985). *Intrinsic motivation and self-determination in human behavior*. New York, NY: Plenum Press.
- Defense Language Institute Foreign Language Center. (2009). *Defense Language Proficiency Testing System 5 Framework*. Retrieved from <http://bit.ly/1cwQwFn>
- de Jong, R., Westerhof, K. J., & Creemers, B. P. M. (2000). Homework and student math achievement in junior high schools. *Educational Research and Evaluation*, *6*, 130–157. doi:10.1076/1380-3611(200006)6:2;1-E;F130
- Duterte, A. (1999). Officer and enlisted personnel success in foreign language learning in DLIFLC's 25-week Spanish basic program. *Dialog on Language Instruction*, *13*, 117–130.
- Eckman, F. R., Iverson, G. K., Fox, R. A., Jacewicz, E., & Lee, S. A. (2011). Explicit training and implicit learning of L2 phonemic contrasts. In C. Sanz & R. P. Leow (Eds.), *Implicit and explicit language learning: Conditions, processes, and knowledge in SLA and bilingualism* (pp. 159–174). Washington, DC: Georgetown University Press.

- Field, A., Miles, J., & Field, Z. (2012). *Discovering statistics using R*. Washington, DC: Sage.
- Flege, J. E. (1987). A critical period for learning to pronounce foreign languages? *Applied Linguistics*, 8, 162–177. doi:10.1093/applin/8.2.162
- Flege, J. E. (1991). Age of learning affects the authenticity of voice onset time (VOT) of stop consonants produced in a second language. *Journal of the Acoustical Society of America*, 89, 395–411. doi:10.1121/1.400473
- Flege, J. E., Munro, M. J., & MacKay, I. R. A. (1995). Factors affecting strength of perceived foreign accent in a second language. *Journal of the Acoustical Society of America*, 97, 3125–3134. doi:10.1121/1.413041
- Flege, J. E., Yeni-Komshian, G. H., & Liu, S. (1999). Age constraints on second-language acquisition. *Journal of Memory and Language*, 41, 78–104. doi:10.1006/jmla.1999.2638
- Foreign Service Institute. (1973). Expected levels of absolute speaking proficiency in languages taught at the Foreign Service Institute. Retrieved from <http://bit.ly/l4cOuM>
- Fox, J. (1997). *Applied regression analysis, linear models, and related methods*. Washington, DC: Sage.
- Friedenreich, C. M., Courneya, K. S., Neilson, H. K., Matthews, C. E., Willis, G., Irwin, M., . . . Ballard-Barbash, R. (2006). Reliability and validity of the Past Year Total Physical Activity Questionnaire. *American Journal of Epidemiology*, 163, 959–970. doi:10.1093/aje/kwj112
- Gardner, R. C., & Lambert, W. E. (1972). *Attitudes and motivation in second language learning*. Raleigh, MA: Newbury House.
- Gass, S. M., Mackey, A., & Pica, T. (1998). The role of input and interaction in second language acquisition: Introduction to the special issue. *Modern Language Journal*, 82, 299–307. doi:10.1111/j.1540-4781.1998.tb01206.x
- Ginns, P., & Barrie, S. (2004). Reliability of single-item ratings of quality in higher education: A replication. *Psychological Reports*, 95, 1023–1030. doi:10.2466/pr0.95.3.1023-1030
- Grasgreen, A. (2013). Languages for all? *Inside Higher Ed*. Retrieved from <http://www.insidehighered.com/>
- Hayden-Wade, H. A., Coleman, K. J., Sallis, J. F., & Armstrong, C. (2003). Validation of the telephone and in-person interview versions of the 7-day PAR. *Medicine & Science in Sports and Exercise*, 35, 801–809. doi:10.1249/01.MSS.0000064941.43869.4E
- Helmore, E. (2013). Policy needs to change to address the US and UK's language deficits. *The Guardian*. Retrieved from www.theguardian.com/
- Henning, G. (1992). The ACTFL Oral Proficiency Interview: Validity evidence. *System*, 20, 365–372. doi:10.1016/0346-251X(92)90046-6
- Herzog, M. (n.d.). *An overview of the history of ILR language proficiency skill level descriptions and scale*. Retrieved from <http://www.govtilr.org/Skills/>
- Hosmer, D. W., & Lemeshow, S. (2000). *Applied logistic regression* (2nd ed.). New York, NY: Wiley.
- Huang, B. H. (2013). The effects of age on second language grammar and speech production. *Journal of Psycholinguistic Research*. Advance online publication. doi:10.1007/s10936-013-9261-7
- Huttenlocher, J., Haight, W., Bryk, A., Seltzer, M., & Lyons, T. (1991). Early vocabulary growth: Relation to language input and gender. *Developmental Psychology*, 27, 236–248. doi:10.1037/0012-1649.27.2.236
- Hyde, J. S., & Linn, M. C. (1988). Gender differences in verbal ability: A meta-analysis. *Psychological Bulletin*, 104, 53–69. doi:10.1037/0033-2909.104.1.53
- Johnson, J. S., & Newport, E. L. (1989). Critical period effects in second language learning: The influence of maturational state on the acquisition of English as a second language. *Cognitive Psychology*, 21, 60–99. doi:10.1016/0010-0285(89)90003-0
- Junco, R. (2013). Comparing actual and self-reported measures of Facebook use. *Computers in Human Behavior*, 29, 626–631. doi:10.1016/j.chb.2012.11.007
- Keith, T. Z., & Cool, V. A. (1992). Testing models of school learning: Effects of quality instruction, motivation, academic coursework, and homework on academic achievement. *School Psychology Quarterly*, 7, 207–226. doi:10.1037/h0088260
- Kimura, D. (1999). *Sex and cognition*. Cambridge, MA: MIT Press.
- Kobayashi, Y. (2002). The role of gender in foreign language learning attitudes: Japanese female students' attitudes towards English learning. *Gender and Education*, 14, 181–197. doi:10.1080/09540250220133021
- Kuhl, P. K., Tsao, F.-M., & Liu, H.-M. (2003). Foreign-language experience in infancy: Effects of short-term exposure and social interaction on phonetic learning. *Proceedings of the National Academy of Sciences, USA*, 100, 9096–9101. doi:10.1073/pnas.1532872100
- Lessells, C. M., & Boag, P. T. (1987). Unrepeatable repeatabilities: A common mistake. *The Auk*, 104, 116–121. doi:10.2307/4087240
- Lett, J. A., & O'Mara, F. E. (1990). Predictors of success in an intensive foreign language learning context: Correlates of language learning at the Defense Language Institute Foreign Language Center. In T. Parry & C. W. Stansfield (Eds.), *Language aptitude reconsidered* (pp. 222–260). Englewood Cliffs, NJ: Prentice Hall Regents.
- 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. doi:10.1111/lang.12011
- Long, M. H. (1981). Input, interaction, and second-language acquisition. In H. Winitz (Ed.), *Annals of the New York Academy of Sciences: Vol. 379. Native language and foreign language acquisition* (pp. 259–278). doi:10.1111/j.1749-6632.1981.tb42014.x
- Miles, D. (2005). *New test to better assess language proficiency within DoD*. Retrieved from <http://1.usa.gov/1a9Nc11>
- Nakagawa, S., & Schielzeth, H. (2013). A general and simple method for obtaining R^2 from generalized linear mixed-effects models. *Methods in Ecology and Evolution*, 4, 133–142. doi:10.1111/j.2041-210x.2012.00261.x
- National Virtual Translation Center. (2007). *Language learning difficulty for English speakers*. Retrieved from <http://bit.ly/11j7X2X>
- Oxford, R. L. (1993). Instructional implications of gender differences in second/foreign language (L2) learning styles and strategies. *Applied Language Learning*, 4, 65–94.
- Patall, E. A., Cooper, H., & Robinson, J. C. (2008). The effects of choice on intrinsic motivation and related outcomes: A meta-analysis of research findings. *Psychological Bulletin*, 134, 270–300. doi:10.1037/0033-2909.134.2.270
- Patall, E. A., Cooper, H., & Wynn, S. R. (2010). The effectiveness and relative importance of choice in the classroom. *Journal of Educational Psychology*, 102, 896–915. doi:10.1037/a0019545
- Patkowski, M. S. (1990). Age and accent in a second language: A reply to James Emil Flege. *Applied Linguistics*, 11, 73–89. doi:10.1093/applin/11.1.73
- Petersen, C. R., & Al-Haik, A. R. (1976). The development of the Defense Language Aptitude Battery (DLAB). *Educational and Psychological Measurement*, 36, 369–380. doi:10.1177/001316447603600216
- Pufahl, I., Rhodes, N. C., & Christian, D. (2000). *Foreign language teaching: What the United States can learn from other countries*. Washington, DC: Center for Applied Linguistics.
- Pufahl, I., Rhodes, N. C., & Christian, D. (2001). *What we can learn from foreign language teaching in other countries*. Washington, DC: Center for Applied Linguistics.
- Raudenbush, S. W., & Bryk, A. S. (2002). *Hierarchical linear models: Applications and data analysis methods* (2nd ed.). Thousand Oaks, CA: Sage.
- R Development Core Team. (2013). *R: A language and environment for statistical computing*. Available from <http://www.r-project.org>

- Rhodes, N. C., & Pufahl, I. (2009). *Foreign language teaching in U.S. schools: Results of a national survey*. Washington, DC: Center for Applied Linguistics.
- Rodgers, D. (2013). *The best DLAB study guide: Practice test and more*. Available from <http://www.amazon.com/AmazonDigitalServices>
- Ryan, R. M., & Deci, E. L. (2000). Self-determination theory and the facilitation of intrinsic motivation, social development, and well-being. *American Psychologist*, *55*, 68–78. doi:10.1037/0003-066X.55.1.68
- Sallis, J. F. (1991). Self-report measures of children's physical activity. *Journal of School Health*, *61*, 215–219. doi:10.1111/j.1746-1561.1991.tb06017.x
- Schacht, S. P., & Aspelmeier, J. E. (2005). *Social and behavioral statistics: A user-friendly approach* (2nd ed.). Boulder, CO: Perseus Books.
- Schumann, J. H. (1986). Research on the acculturation model for second language acquisition. *Journal of Multilingual and Multicultural Development*, *7*, 379–392. doi:10.1080/01434632.1986.9994254
- Seliger, H. W. (1977). Does practice make perfect? A study of interaction patterns and L2 competence. *Language Learning*, *27*, 263–278. doi:10.1111/j.1467-1770.1977.tb00122.x
- Skorton, D., & Altschuler, G. (2012). America's foreign language deficit. *Forbes*. Retrieved from <http://onforb.es/U5pq6>
- Smith, T. E. (1990). Time and academic achievement. *Journal of Youth and Adolescence*, *19*, 539–558. doi:10.1007/BF01537175
- Smith, T. E. (1992). Time use and change in academic achievement: A longitudinal follow-up. *Journal of Youth and Adolescence*, *21*, 725–747. doi:10.1007/BF01538741
- Snijders, T. A. B., & Bosker, R. J. (2012). *Multilevel analysis: An introduction to basic and advanced multilevel modeling* (2nd ed.). London, United Kingdom: Sage.
- Stølen, M. (1987). The effect of affect on interlanguage phonology. In G. Ioup & S. H. Weinberger (Eds.), *Interlanguage phonology: The acquisition of a second language sound system* (pp. 389–400). Cambridge, MA: Newbury House.
- Strath, S. J., Bassett, D. R., Jr., Ham, S. A., & Swartz, A. M. (2003). Assessment of physical activity by telephone interview versus objective monitoring. *Medicine & Science in Sports and Exercise*, *35*, 2112–2118. doi:10.1249/01.MSS.0000099091.38917.76
- Surface, E. A., & Dierdorff, E. C. (2003). Reliability and the ACTFL Oral Proficiency Interview: Reporting indices of interrater consistency and agreement for 19 languages. *Foreign Language Annals*, *36*, 507–519. doi:10.1111/j.1944-9720.2003.tb02140.x
- Tercanlioglu, L. (2004). Exploring gender effect on adult foreign language learning strategies. *Issues in Educational Research*, *14*, 181–193.
- Walker, D. A. (2003). Converting Kendall's tau for correlational or meta-analytic analyses. *Journal of Modern Applied Statistical Methods*, *2*, 525–530.
- Wallinger, L. M. (2000). The role of homework in foreign language learning. *Foreign Language Annals*, *33*, 483–496. doi:10.1111/j.1944-9720.2000.tb01993.x
- Watanabe, Y., & Swain, M. (2007). Effects of proficiency differences and patterns of pair interaction on second language learning: Collaborative dialogue between adult ESL learners. *Language Teaching Research*, *11*, 121–142. doi:10.1177/136216880607074599
- Wells, G. (1981). *Learning through interaction: The study of language development*. Cambridge, United Kingdom: Cambridge University Press.
- Williams, J. N. (2009). Implicit learning in second language acquisition. In W. C. Ritchie & T. K. Bhatia (Eds.), *The new handbook of second language acquisition* (2nd ed., pp. 319–353). Bingley, United Kingdom: Emerald Group.
- Wode, H. (1994). Nature, nurture, and age in language acquisition: The case of speech perception. *Studies in Second Language Acquisition*, *16*, 325–345. doi:10.1017/S0272263100013115

Received January 11, 2013

Revision received January 15, 2014

Accepted February 11, 2014 ■