Language Learning and Metacognition: An Intervention to Improve Language Classrooms

Rebecca E. Knoph

University of Oslo

Erin M. Buchanan

Harrisburg University of Science and Technology

Abstract

In the USA, the trend of increase in foreign language enrollments at the college level has suddenly begun to decline since 2009, despite the notion that learning multiple languages is becoming essential for effectively communicating with others from diverse native language backgrounds. This new decline may be due in part to inefficient and outdated foreign language courses. The current study examined the effect of how we assess our current knowledge and learning techniques (metacognition) on educational outcomes in hopes to improve the effectiveness of the university classrooms. College students were exposed to new metacognitive strategies that could benefit their language learning throughout the fall 2016 semester. Specifically, students were presented with new information every other week to improve their vocabulary building, listening skills, and writing skills. Hierarchical multiple linear regression provided evidence of a potential way to measure and promote metacognitive strategies that could be useful to language learners.

Keywords: language learning, metacognition, motivation, academic achievement, structured learning

Although the value of learning multiple languages affects many Americans, the current education system in the USA does not generally support multilingual learning from an early age for native speakers of English. Instead, most foreign language learning occurs later in adult life and in structured classrooms, which may not be efficient in long-term learning. The current study aimed to evaluate the effectiveness of learning in university classrooms when combining language learning and metacognitive awareness for native speakers of English who were learning different target languages.

Metacognition

Commonly, metacognition is known as "thinking about thinking", although John Flavell coined the term as "cognition about cognitive phenomena" (Flavell, 1979, p. 906). Within the educational field, O'Malley and Chamot (1990) have since expanded the definition to include thinking, planning, and self-evaluation of one's learning.

Metacognition Instruction. Most metacognitive learning requires informal strategies through practice (Schraw, Crippen, & Hartley, 2006), but some strategies may be learned (and thus taught) through interventions. For example, Cross and Paris (1988) discussed an intervention for 171 elementary students to improve reading comprehension. Students were taught new metacognitive reading strategies focusing on declarative knowledge ("*what* factors influence reading"), procedural knowledge ("*how* skills operate and are applied"), and some cognitive planning/conditional knowledge ("understanding *when* particular strategies should be applied, and why they affect

Correspondence concerning this article should be addressed to Rebecca Knoph, Department of Education, University of Oslo, Postboks 1092, Blindern, 0317 OSLO.

reading"). Students who received strategy instruction performed significantly better than the control group, especially in the areas of task difficulty evaluation, goal planning, and goal monitoring (Cross & Paris, 1988).

Moreover, Kramarski and Mevarech (2003) discussed the effects of metacognitive instruction of over 380 teenage students, specifically in their mathematical skills and reasoning. Students receiving metacognitive instruction significantly outperformed their peers in a multitude of ways, such as interpreting graphs, using logic to support arguments, and transferring mathematical logic from one concept to another. Similarly, Wolters (2004) found that, for seventh and eighth grade students enrolled in a variety of mathematics courses, practicing study strategies was correlated more strongly with final grade than using cognitive strategies (r = .21 and r = .11, respectively). Additionally, Vrugt and Oort (2008) found a moderate relationship between use of general metacognitive strategies and introductory psychology exam scores (β = .20) when assessing structural equation modelling for 800 undergraduates.

Likewise, several meta-analyses support the effectiveness of teaching metacognition to improve student learning with moderate effect sizes (Dignath, Buettner, & Langfeldt, 2008; Haller, Child, & Walberg, 1988; Raoofi, Chan, Mukundan, & Rashid, 2013). For example, a meta-analysis comprised of 20 studies and over 1,500 students conducted by Haller et al. (1988) proposed that metacognitive instruction can have immense effect on reading comprehension for children of varying ages (mean effect size d = 0.71). They suggested techniques such as: teaching awareness of inconsistency in reading, using self-questioning to monitor comprehension, and using self-reflection to regulate speed of reading.

Dignath et al. (2008) conducted a meta-analysis comparing metacognition and learning outcomes in all disciplines for children in first through sixth grade. The mean effect size for the 48 studies was d = 0.73; again, a sizable effect which showed that children can learn metacognitive strategies and those strategies can have a positive impact on a variety of learning outcomes. A more recent meta-analysis conducted by Raoofi et al. (2013) came to similar conclusions, although no effect size was reported as the analyzed studies were qualitative. Here, the selected studies focused on interviewing young students and their opinions of using metacognition, or asking students to think aloud while researchers recorded their thoughts. Similar to the quantitative results of previous researchers, Raoofi et al. (2013) found that children enjoyed using metacognitive strategies and found their learning to be more efficient and long lasting in most educational areas than when they did not use metacognitive strategies.

While empirical evidence suggests that metacognitive instruction can provide benefit to learners in most areas, it appears that teachers are relatively unaware of metacognition as a concept (Veenman, Van Hout-Wolters, & Afflerbach, 2006). Further, the majority of teachers who were aware of the concept of metacognition did not feel confident in their ability to instruct students on metacognitive strategies (Veenman et al., 2006). Yet, most researchers agree that instruction in metacognitive strategies is not only possible, but also preferred for metacognitive instruction. Schraw et al. (2006) suggested that metacognitive strategies should be presented in explicit form to students, and then followed up with an explanation as to why a strategy is beneficial. Providing an explanation may increase motivation and thus independent practice with the strategy (Cross & Paris, 1988; Kramarski & Mevarech, 2003). Therefore, the current study focused on exposure to explicit strategies that participants could learn and improve upon, along with benefits for using the new strategies. One-week breaks between intervention exposure allowed for participants to practice putting strategies into effect in the most efficient way for each participant's individual learning preferences, target language, and course level.

Limited research has focused on the use of metacognitive strategies specifically within the language-learning domain as a whole, and especially of young adults who are learning a language at university. Based on the meta-analysis conducted by Raoofi et al. (2013), only one study had considered this specific group (albeit captured in a wider range of ages): Pishghadam and Khajavy (2013), who studied 143 Persian students from ages 10 – 40 learning English. After measuring participants' intelligence and metacognitive awareness on the Metacognition Awareness Inventory (MAI), the researchers found that scores on the MAI could account for 17% of the variance in standardized grades after controlling for intelligence. However, this study did not control for age, despite the wide range of the sample. The current study aimed to shed light on how native speakers of English might benefit from metacognitive strategies when studying different foreign languages offered at the university level.

Metacognitive Strategies within Language Learning

Although metacognition and second language learning research remains mostly preliminary, some strategies used to learn our native language (NL) could theoretically be transferred to learning a new target language (TL). For instance, learners could predict the next word in a sentence, use contextual and paralinguistic cues to guess meaning, or evaluate their comprehension and ask for clarification when necessary (August, Calderon, & Carlo,

2002; 1990; Oxford, 1990; Wang, Thomas, & Ouellette, 1992). Further, metacognition is a vital part of language learning, as otherwise learners cannot review their progress and future learning (O'Malley et al., 1985).

O'Malley et al. (1985) assessed the learning strategies used by successful TL learners and found that some of the most frequent strategies involved metacognition - specifically vocabulary strategies, attentional strategies, and planning strategies, which can be used for all four language learning skills: reading, writing, speaking, and listening. Moreover, Moir and Nation (2002) found that learners of English who were more aware of their own English abilities (and their own pitfalls) performed significantly better on vocabulary tests than those that were not as metacognitively aware. The current study provided direction and planning as a supplement to the cognitive learning strategies commonly taught in the classroom and focused on three domains in relation to TL courses: vocabulary, listening, and writing.

Vocabulary Strategies. Students recognize that learning more vocabulary words is imperative for their TL growth; thus, vocabulary-increasing strategies are in high demand (Nation & Webb, 2011; 1990). One such strategy, the keyword method, is a well-researched mnemonic strategy that aids in immediate recall of vocabulary words. In this strategy, a student encounters a new word in the TL (for instance, *vegg*, the Norwegian word for *wall*) and then creates a keyword: a word that sounds or looks similar to the novel word (for instance, *egg*). Then, the student interacts the keyword (*egg*) with a mental image to facilitate remembering the meaning of the novel word (for instance, thinking of someone throwing eggs at a wall; Atkinson, 1975; Kulhavy & Swenson, 1975; Pressley, Levin, & Delaney, 1982). O'Malley and Chamot (1990) found that mnemonic strategies such as the keyword method allow learners to use an item's embedded meaningfulness, such as connecting *vegg* with *egg* and *wall*. Mnemonic strategies can be more beneficial than simply trying to recall the information or using rote memorization (O'Malley & Chamot, 1990; Nation & Webb, 2011). O'Malley and Chamot (1990) confirmed findings from Kulhavy and Swenson (1975) that participants using the keyword method were able to read a passage and provide missing words from paraphrased sentences better than control participants.

However, empirical evidence for the longitudinal effects of the keyword method is mixed. Wang et al. (1992) discussed that the keyword method is beneficial initially, but that the effects diminish over time and are less effective when learning abstract words (e.g., *honesty*), as opposed to concrete words (e.g., *church*). Moreover, Wang et al. (1992) examined if the long-term effects of the keyword method might be a testing effect. When studying native speakers of English who were learning Chinese, the authors found that college students who used the keyword method outperformed their peers on an immediate recall test, but that college students who took only the two-week delayed recall test did not perform significantly different from their control-group peers (Wang et al., 1992).

Avila and Sadoski (1996) endeavored to replicate the previous study, but with 11-year-old native speakers of Spanish who learned English target words. Participants who used the keyword method did perform significantly better than their control group counterparts regardless of test (immediate versus delayed test; main effect of study strategy), and participants taking the immediate test performed better than participants taking the delayed test regardless of strategy (main effect for test), but no significant interaction was present. The effect sizes for post-hoc tests that compared participants who used the keyword method to the control condition were $d = 0.59$ initially and $d = 2.27$ on the delayed recall test (Avila & Sadoski, 1996). Due to mixed results for using the keyword method over long periods of time, participants in the current study were advised to use the keyword method initially and gradually move toward committing novel words to memory via context and lexical inferencing.

Fraser (1999) interviewed adults learning English as a foreign language and discovered that, when participants encountered novel English words, the learners relied solely on contextual clues and lexical inferencing 58% of the time. However, multiple researchers have urged that using context is only effective when the reader knows enough vocabulary from the TL to understand the context (i.e., at least 5,000 words in the TL; Coady, Magoto, Hubbard, Graney, & Mokhtari, 1993; Laufer, 1997; Nation & Webb, 2011) and enough vocabulary from the context itself (i.e., comprehending fewer than 95% of the words in a given passage will result in general comprehension failure and inability to use context to correctly make predictions; Hirsh & Nation, 1992).

Nassaji (2003) assessed 21 learners of English from a variety of native languages and found that generally, readers were unable to correctly guess the meaning of novel English words, even when they understood 95% of the words in the passage. Participants correctly inferred meaning of new words only 25.6% of the time - significantly less than chance. Usually, participants incorrectly assumed nonexistent semantic relationships, such as assuming *permeate* is related to *meat* (Fukkink, Blok, & de Glopper, 2001; Morrison, 1996; Nassaji, 2003; Schatz & Baldwin, 1986). Huckin and Bloch (1993) proposed that making contextual guesses is difficult because it requires two components: first, readers must generate different possible meanings based on known words that behave similarly to the new word. For instance, if readers are presented with the nonword *donse* in the sentence, "Please *donse* the flowers to Rose", readers must first consider multiple potential meanings (e.g., *give, present, roll…*) based on the

sentence grammar. Next, readers must evaluate which hypothesized meaning is most likely correct based on lexical semantics. Since flowers are not usually rolled, the readers would want to exclude the hypothesis that *donse* means something similar to roll and instead conclude that *donse* probably means something similar to *give* or *present*.

Despite methodological shortcomings of vocabulary methods, the keyword method has been extensively investigated and established as an effective tool for initial recall of new TL words (Nation & Webb, 2011). When combined with the tools to use lexical inferencing for abstract words over time, we expected to see participants able to use these strategies effectively (Fukkink et al., 2001; Moore & Surber, 1992).

Listening Strategies. Goh (2008) expanded on three metacognitive strategy categorizations for TL listening comprehension first suggested by Brown (1978): planning, monitoring, and evaluating. Goh (2008) suggested that during the *planning* phase, listeners should set goals and seek opportunities to practice listening. Listening is not limited to conversing with natives, but should include authentic materials, such as television or radio shows. Authentic materials are not created with a second language learner in mind, but they are instead created within the domain of the TL itself and tend to have faster speech that has less emphasis on enunciation (Gilmore, 2007). Goh (2008) suggested that participants seek out these materials and, if needed, use subtitles as a crutch. During the *monitoring* phase, listeners should check for accurate comprehension while listening. This can include asking questions during a conversation or by rewinding shows. Listeners should also consider if it is realistic to meet the goals set during the planning phase - or if they need to adjust their listening strategies or goals. Finally, listeners should *evaluate* their overall understanding at the conclusion of the listening task by drawing on their background knowledge and considering if what they believe they heard sounds logical (Goh, 2008). Listeners should also consider if they met their goals by the end of the task, evaluate why this accomplishment did or did not happen, and how to make it happen in future tasks. Participants in the current study were exposed to these same tools.

Writing Strategies. Schoonen et al. (2003) expanded on a similar metacognitive strategy categorization for TL written production first suggested by Chenoweth and Hayes (2001): resource, process, and control. For Schoonen et al. (2003), *resource* refers to the metacognitive knowledge the writer currently holds, such as knowing how to: organize writing thoughts, consider outside resources available when writing, and decide which writing strategies are best for the current task. *Process* refers to accessing those available resources, which is constrained within the TL and is limited by working memory. This step takes a rough outline or a brainstorm of ideas and converts it into a rough draft. *Control* refers to putting the writing together in an organized manner. This step transforms a rough draft into a final, cohesive text.

Most importantly, Schoonen et al. (2003) emphasized the need for writers to consider the thoughts and ideas they wish to communicate to others before considering the restraints of TL ability. Often, Schoonen et al. (2003) argued, second language learners begin with the words and phrases they know they can use correctly and then move towards what ideas they can communicate with only those words and phrases - essentially avoiding any errors in their writing but limiting their expression. This procedure is not what usually happens when writing in a NL, where writers usually begin with what ideas they want to express and then find ways to express those ideas. This process may result in a greater likelihood for errors, but also a larger pool of thoughts to express.

Additionally, researchers suggest that learners practice writing as much as possible. Robertson (2010) specifically suggests that writing short cinquain poems can provide learners with writing practice that takes little time. Practicing with short texts provide learners with bursts of writing that can be completed in any setting, as opposed to requiring a few hours to contemplate. Practice can improve working memory limitations and general TL knowledge.

Hypotheses

We theorized that by combining some of the common metacognitive strategies into a time-series design, we could see trends in learning outcomes. Specifically, we hoped that university students enrolled in the initial courses for a modern language would have higher final course grades after being presented with metacognitive strategies. Additionally, we predicted that these same students would be more likely to continue to the next TL course. Two hypotheses were examined in the current study to determine the effectiveness of the proposed interventions explained previously[1]:

1. Participation in the study and exposure to the interventions, the type of intervention, and their potential interaction would positively predict final course grades.
2. Participation in the study and exposure to the interventions, the type of intervention, and their potential interaction would positively predict persistence to enroll in the next course.

[1] Two other hypotheses were examined, and all results from these can be found at https://osf.io/wfbqd/.

Method

Participants

After receiving IRB approval, participants were recruited through 18 semester-long language courses offered at a large Midwestern university from nine modern languages (Arabic, Chinese, French, German, Italian, Japanese, Portuguese, Russian, and Spanish) and two course levels (the introductory course "101" and the second year basics course "201"). Participants studying Greek, Hebrew, and Latin courses were omitted because these language courses do not function similarly to, nor have similar goals as, the modern language courses.

Of the 547 unique students enrolled in one of the 18 language courses, 32 students initially volunteered to participate in the study. Because of the small sample size, a second wave of recruitment occurred during the midpoint of the study. The second wave of students who volunteered increased the participant pool by 11 (see Table 1). Thus, 43 total students participated in the study with a mean age of 20.35 (SD = 4.70) in which 14% were male and the average number of sessions participated in was 2.26 sessions (M = 2.47 out of 5 sessions for first wave students and M = 1.64 out of 2 sessions for second wave students). Participation was voluntary, and participants were entered into drawings for gift cards as compensation. You can view the exact recruitment flyer used in the study at: https://osf.io/wfbqd/. Participants were identified based on student number (for demographic information such as final grade, GPA, etc.) and on preferred email (for information assessed during sessions). Although data did not remain anonymous, it remained confidential.

Table 1

Participant course tabulation by recruitment wave.

	First Wave		Second Wave	
Target Language	Introductory Course	Advanced Course	Introductory Course	Advanced Course
Spanish	7	5	4	1
French	5	4	1	1
German	2	2	1	0
Arabic	1	0	0	0
Chinese	0	3	0	1
Italian	2	0	1	0
Portuguese	1	0	0	0
Russian	0	0	0	1

Note. This table represents the number of students who participated in the study (N = 43). Randomly selected nonparticipants were matched based on target language and course level, thus the current study includes exactly double the values shown (N = 86; 43 participants and 43 matched nonparticipants).

The study compared students who participated in the study (experimental group) to matched students who chose to opt out of the study (control group). Students in the control group were randomly selected from the remaining population of students enrolled in the same target language and course number (i.e. a participant taking German 201 was matched with a student who was also taking German 201; no other matching variables were considered). Still, the matched nonparticipant sample was not significantly different ($\alpha < .01$) from the participant sample based on age (M = 22.02, SD = 5.48, Welch's $t(82.10) = 1.52$, $p = .13$), nor GPA (M = 3.05, SD = 0.60, $t(84) = 2.24$, $p = .03$), but did contain significantly more males (40% male, $\chi^2(1) = 8.46$, $p < .001$). Regardless, statistical analyses controlled for each of these variables.

Languages in which fewer than five students participated in the study were then combined into an "other" category for data analysis, such as Russian and Chinese. The final sample size for each target language (TL) and course level, and the demographics for sample by time of recruitment is included in the online supplement at https://osf.io/wfbqd/.

Materials

Demographics Questionnaire. During the initial session, participants received a demographics questionnaire to identify the TL and course level, and to allow for statistical control of characteristics such as previous experience with the TL, enrollment in multiple TLs, age, gender, and identifying the TL. After identifying the TL and course level, later items in the initial session "piped" course information. That is, later items referring to TL included each participant's own TL (e.g. "I try to use French frequently") instead of a general placeholder (e.g. "I try to use [target language] frequently"). The basic demographics were assessed only during the initial session, along with some demographic-type questions during the final session, such as "What grade do you predict to receive in this class?"

Motivation Questionnaire. Because the motivation to learn a language is distinct from the motivation to complete other academic tasks, a motivation questionnaire specifically designed for the context of language learning was employed. Wimolmas (2013) adapted Gardner's Attitude and Motivation Test Battery (Gardner, 2004) to function in the context of language learning for Arabic students in their initial year of undergraduate programs learning English, and this questionnaire has been adapted to university students learning other TLs. This language learning questionnaire assesses overall language learning motivation using the two factors discussed earlier by Gardner (2007): instrumental and integrative motivation. The scale contains 10 items for each factor along with one final open-ended question. Scores on each item range from 1-5 with "1" being low motivation and "5" being high motivation, and an overall summed score for general motivation being calculated at the conclusion. The reliability for this scale in our sample was Cronbach's $\alpha = .82$.

Strategy Inventory for Language Learners. The Strategy Inventory for Language Learners (SILL, version 7.0) contains a total of 50 questions that assess what strategies were used to learn, and to what degree each strategy was used in the learning process, specifically when learning a foreign language (Oxford, 1990). The SILL is one of the most commonly used inventories for the measurement of language learning strategy use. The SILL has repeatedly shown adequate to excellent levels of internal reliability, with Cronbach's alpha ranging from .67 to .96 for the English version (Hong-Nam & Leavell, 2006; Hsiao & Oxford, 2002; Nyikos & Oxford, 1993; Robson & Midorikawa, 2001).

Because the inventory is so lengthy, participants were not asked to complete the SILL in its entirety during any one session. Instead, the relevant items on the SILL for each intervention were assessed in the following session. That is, since Intervention 2 presents strategies for improving vocabulary, SILL items such as "I use new target language words in a sentence so I can remember them" were asked during the next session (Session 3). Asking during the next session provided participants with actual time to practice using the strategies presented before being asked if they did or did not use said strategies. The specific items from the SILL assessed during each session (as they relate to the previous intervention) are found at https://osf.io/wfbqd/.

Interventions. The study contained five sessions for participation, but only exposure to four interventions (exposure to modules for improving metacognition). Each intervention exposed participants to a text transcription of explicit metacognitive strategies and an online video with the text presented while being read aloud. Both mediums allowed participants to read at his or her own pace, take notes, or read/listen again. The presentation of new strategies for each intervention took approximately 10 minutes to read or to listen. Interventions 2-4 follow the information given in the introduction. All intervention material is available online.

Intervention 1: Basic Metacognition did not include any explicit tips on how to improve metacognition, but rather asserted the basics of general metacognition and what to expect throughout the study, as many students are unaware of what metacognition is (Karpicke, 2009).

Intervention 2: Improving Vocabulary exposed participants to tips on improving vocabulary via the keyword method and semantic context method described earlier. We proposed that by introducing an especially effective method relevant to all learning stages, we could reduce the potential attrition in the current study due to loss of interest in the presented strategies.

Intervention 3: Successful Listening exposed participants to tips on improving listening and information about linguistic motivation. Listening is usually viewed as the most daunting task for language comprehension, as evidenced by the current participants (46% answered Listening as the hardest part of learning a language) and previous studies (Vandergrift, 2004; Wimolmas, 2013). Information about linguistic motivation was also provided during this specific session because the enrollment period for the next semester began around the time of session 3 deployment. It was key for participants to understand how they could feel motivated to learn and be able to consider how that might affect their future enrollment decisions before having to make said decisions. Thus, the sub-intervention of motivation was added to the listening intervention.

Intervention 4: Improving Writing encompassed another difficult linguistic task: production of written language. Although speaking is also found to be daunting, students in the introductory course (101) may not reach a point where speaking novel utterances in the TL is required. Because the relevance of speaking tips could not be guaranteed for a majority of participants, writing tips were deemed most potentially beneficial for all.

Procedure

Recruitment. First, an e-mail was sent to approximately 550 students enrolled in a modern language 101 or 201 course with a link to register their phone number using Remind 101 and a note that participation was voluntary. Using Remind 101 allowed participants' information to remain private; no investigators were able to access provided phone numbers.

Midway through the study, a second recruitment e-mail was sent to encourage students who had not yet registered to participate in the study. This second wave of participants aimed to aid the retention rate and small sample size. Students who volunteered to participate during the second recruitment phase were assigned to Session 1 while the original participants were assigned to Session 3. Afterwards, both groups of participants were assigned to Session 4 and later Session 5.

Sessions. Once recruited, students received a text with the link to begin the study. The link directed participants to the first online session comprised of the consent form, basic demographics questionnaire, and the motivation questionnaire, in addition to the first intervention, manipulation check, and attitudes toward the first intervention. The session remained open for one week (Monday through Sunday). Participants who did not complete the session before the midway point of session deployment received a second text via Remind 101. Participants who did not complete the session before the final day of session deployment received a final reminder text. After each session was open for one week, participants were asked to use the following week to practice the strategies presented during the session. During this time, no new sessions were made available. Thus, sessions occurred on alternating weeks and participants could choose when in the week they wanted to participate.

After the week to practice, a new text was sent to all participants urging to participate in the next session along with a new link. The new link redirected to the new session, which included a new consent form, the next intervention, and the manipulation check for said intervention. Information on participant accuracy for each manipulation check is included online in the supplemental material. Once completing interventions and manipulation checks, participants were asked about their attitudes regarding the information presented. Participants first answered questions pertaining to the current intervention, such as "Do you feel the intervention presented today will be beneficial?" Next, participants answered questions pertaining to their use of the previous intervention, such as "Did you use the tips provided during the *previous* session?" Last, participants answered the relevant items from the Strategy Inventory for Language Learners (SILL) as they relate to the previous intervention. This procedure continued until all five sessions had been announced.

Further, the first and last sessions contained different information from the middle sessions. Session 1 included the demographic questionnaire and the motivation questionnaire, along with the first intervention. It also did not include any information about previous interventions as none had occurred yet. Session 5 did not contain any new intervention material, but instead participants were given the survey for attitudes about Intervention 4 and relevant SILL questions from Intervention 4, a final survey about the study as a whole, and a final motivation questionnaire.

Design. A repeated measures time series design was employed for the intervention part of the study and included four measurement points, one for each intervention described earlier. For data analysis, all information pertaining to an intervention was included as one-time point, despite some information being collected later in the study. That is, even though items assessing if participants used the vocabulary strategies (Session 3) were employed later than all other questions assessing vocabulary strategies (Session 2), all information collected about vocabulary strategies was considered one measurement point: Intervention 2 information. After the semester was completed, documented data for final course grade, GPA (including the current semester), and continued enrollment in the TL was collected.

Results

Data Analysis

Hierarchical multiple linear regression and hierarchical logistic regression were used to answer the overall research questions and control for demographic variables such as GPA, gender, age, TL, and course level. In each hypothesis, we first examined class factors predicting final grade, followed by models that added student

demographics, GPA, and finally our intervention. Using hierarchical regression, we were able to examine the additional variance predicted above and beyond the previous variables (i.e., while controlling for the previous steps). The interaction between type of intervention (control, vocabulary, listening, and writing), and participation in each intervention individually (yes or no) was not analyzed due to small sample sizes. These analyses were conducted using *R*.

Prediction of Final Grade based on Voluntary Participation

First, we predicted that participation in, and exposure to, the metacognition interventions would positively predict final course grade. The data were screened for missing data, outliers, and assumptions for regression. Although two participants were considered outliers (i.e., their data met the cutoff for at least two of the following: Cook's values, Leverage values, and/or Mahalanobis distances), their data were retained in the analyses since a participant cannot truly be an outlier for educational outcomes (i.e., their final course grades were accurate). Additionally, the inclusion of outliers did not change the interpretation of the final inferential statistics. Normality, linearity, homogeneity, homoscedasticity, and additivity were all met.

We predicted that several variables may influence final grade, thus, a hierarchical model was analyzed to test the ability of participation to uniquely predict final grade. First, differences in courses were considered. The first model controlling for TL (using Spanish as the dummy-coded reference group), course level (using Introductory, or 101, for reference), and average course grade for that specific section of the course (to control for potential differences in grading style and inflation) was not significant, $F(4, 80) = 1.80$, $p = .14$, $R^2 = .08$. Thus, TL and course level could not predict final grade. However, the average course grade for that section did predict final grade, $b = 1.00$, $t(80) = 2.52$, $p = .01$, $pr^2 = .07$.

Next, differences in student demographics of age and gender (using males as the reference group) were analyzed. The addition of these variables was significant, $F(2, 78) = 3.99$, $p = .02$, $\Delta R^2 = .05$, Overall $R^2 = .13$. The only significant predictor in this step of the analysis was age, indicating that younger students performed better in their courses, $b = -0.06$, $t(78) = -2.08$, $p = .04$, $pr^2 = .05$.

Then, differences in student GPA were considered. Empirical evidence supports that current and previous GPAs are a significant predictor of future grades (Hodara & Cox, 2016; Hodara & Lewis, 2017), thus, it was no surprise that the addition in this model was significant, $F(1, 77) = 41.60$, $p < .001$, $\Delta R^2 = .28$, Overall $R^2 = .41$. When participants had a higher GPA, they were more likely to obtain a higher final course grade in their target language, $b = 1.02$, $t(77) = 5.96$, $p < .001$, $pr^2 = .32$, Finally, the addition of participation as a predictor for final course grade was significant, $F(1, 76) = 14.01$, $p < .001$, $\Delta R^2 = .09$, Overall $R^2 = .50$. When participants decided to participate in the study and were exposed to metacognitive strategy instruction, they were more likely to achieve a higher final grade, $b = 0.82$, $t(76) = 3.74$, $p < .001$, $pr^2 = .16$. The moderate effect size for pr^2 alludes that exposure to metacognitive strategies can predict final grades. For all predictor information, see Table 2.

A power analysis for hierarchical regression, with $\alpha = .05$ and power = .80, indicated that $N = 82$ participants should have been collected for this the final ΔR^2 (Faul, Erdfelder, Lang, & Buchner, 2007). Additionally, we examined the results for these analyses on first wave participants only, and the direction of the effects remain the same.

Prediction of Persistence based on Voluntary Participation

Next, we predicted that participation in the metacognition interventions would positively predict persistence (enrollment in the next course). Students enrolled in TL 101 who then enrolled in TL 102, and students in TL 201 who enrolled in TL 202 were considered to persist in their target language learning. Likewise, students enrolled in TL 101 who immediately enrolled in TL 201 were considered to persist ($N = 3$). Students who were retaking the current course, taking a course in any other language, or failing to take any language courses were considered to not persist.

A hierarchical binary logistic regression analysis was conducted to evaluate the prediction of persistence using participation after controlling for course and individual differences. The data were screened for missing data, outliers, and assumptions for logistic regression (all were met). Due to prerequisite limitations, all students receiving a grade below C (2.0) were omitted for this analysis ($N = 14$; $N = 1$ participant, $N = 13$ nonparticipants). Of the students who were able to decide to continue enrollment, 44 persisted and 27 did not.

Table 2

Model information for Hypothesis 1 prediction of course grade given participation.

Model	Variable		*b*	*SE*	*t*	*p*	pr^2
1	Target Language						
1		Spanish vs French	-0.16	0.35	-0.47	.64	.002
1		Spanish vs Other	-0.09	0.32	-0.28	.78	.001
1	Course Level		0.05	0.27	0.20	.84	< .001
1	Avg Final Grade		1.00	0.39	2.52	.01	.07
2	Age		-0.06	0.03	-2.08	.04	.05
2	Gender		-0.03	0.33	-0.08	.94	< .001
3	GPA		1.02	0.17	5.96	< .001	.32
4	Participation		0.82	0.22	3.74	< .001	.16

Note. Target language was represented as three dummy variables with Spanish serving as the reference group; course level with introductory serving as the reference group; gender with males serving as the reference group; and nonparticipation serving as the reference group. Values represent the step they were entered into the model.

The first regression step which included TL, course level, and average final grade for that section of the course was not significant, $\chi^2(4) = 4.28$, $p = .37$, Nagelkerke's Pseudo $R^2 = .08$, thus, indicating that course variables alone could not predict persistence in enrollment. Next, individual demographic variables (age and gender) were included in the analysis. The addition of variables to the model was significant, $\chi^2(2) = 6.28$, $p = .04$, Nagelkerke's Pseudo $R^2 = .19$, indicating that the addition of course and demographics predicted persistence. However, neither variable was individually significant, age $b = -0.13$, $Z = -1.82$, $p = .07$ and gender $b = 0.53$, $Z = 0.83$, $p = .41$.

Then, current student GPA was included in the analysis. Surprisingly, this model change was not significant, $\chi^2(1) = 0.26$, $p = .61$, Nagelkerke's Pseudo $R^2 = .19$. Thus, despite previous research indicating that GPA is a significant predictor in retention (Reason, 2003), GPA was unable to provide predictive power in the current study after controlling for previous variables. Finally, participation in the interventions was added to the analysis. This final model also produced nonsignificant results, $\chi^2(1) = 0.50$, $p = .48$, Nagelkerke's Pseudo $R^2 = .20$. Thus, when controlling for course level, TL, average course grade, age, gender, and GPA, participation in and exposure to metacognitive strategies were unable to account for significant unique variance. Overall, 69.0% of participants were correctly predicted in their persistence, with better predictions for those persisting (81.8%) compared to those ending their TL university education (48.2%). For all predictor information, see Table 3. For this analysis, it was estimated that 153 participants would be necessary for power for the final step of the model. Examining first wave participants only did not change the results.

Prediction of Strategy Use for Participants

Finally, an exploratory analysis evaluating the relationship between the mean perceived usefulness of metacognitive interventions and scores on the Strategy Inventory for Language Learning (SILL) was considered. The data were screened for missing data, outliers, and assumptions for regression, Participants that did not indicate their perceived usefulness at least once were omitted from further analysis ($N = 14$). Using previously stated standards, no participants were considered outliers.

Analogous to previous hypotheses, differences in courses were controlled for first. The first model controlling for TL, course level, and average course section grade was not significant, $F(4, 14) = 0.20$, $p = .94$, $R^2 = .05$. Thus, the course alone was unable to predict mean score on relevant SILL items. Individual age and gender (using males as the reference group) were controlled for next. Again, this model was not significant, $F(2, 12) = 1.07$, $p = .38$, $\Delta R^2 = .09$, Overall $R^2 = .14$, indicating that course, age, and gender together could not predict mean SILL score. Then, differences in student GPA were considered and was also nonsignificant, $F(1, 11) = 0.68$, $p = .43$, $\Delta R^2 = .03$, Overall $R^2 = .17$. Next, differences in individual motivation scores were analyzed. Again, motivation scores were not significant, $F(1, 10) = 0.16$, $p = .70$, $\Delta R^2 = .01$, Overall $R^2 = .18$, indicating that motivation itself could not uniquely predict mean SILL score when controlling for GPA and other variables. Finally, the inclusion of mean

perceived usefulness as a predictor for mean SILL score was significant, $F(1, 9) = 10.85$, $p = .001$, $\Delta R^2 = .45$, Overall $R^2 = .63$. When participants found the interventions to be useful, they were more likely to also have a high mean score on the relevant SILL items ($b = 0.69$, $t(9) = 3.30$, $p = .01$, $pr^2 = .55$). The large effect sizes ($\Delta R^2 = 0.45$; $pr^2 = .55$) provide evidence that perceived usefulness of the interventions was a strong predictor of SILL scores, as 55% of the variance not accounted for by TL, course level, age, gender, GPA, and motivation was accounted for by perceived usefulness. For all predictor information, see Table 4.

Table 3

Model information for Hypothesis 2 prediction of persistence given participation.

Model	Variable		*b*	*SE*	*z*	*p*
1	Target Language					
1		Spanish vs French	0.41	0.66	0.62	.53
1		Spanish vs Other	0.37	0.60	0.61	.54
1	Course Level		0.90	0.54	1.68	.09
1	Avg Final Grade		0.16	0.77	0.20	.84
2	Age		-0.13	0.07	-1.82	.07
2	Gender		0.53	0.64	0.83	.41
3	GPA		0.22	0.43	0.51	.61
4	Participation		0.42	0.59	0.71	.48

Note. Target language was represented as three dummy variables with Spanish serving as the reference group; course level with introductory serving as the reference group; gender with males serving as the reference group; nonparticipation serving as the reference group; and no enrollment (non-persistence) serving as the reference group.

Table 4

Model information for exploratory prediction of strategy use given perceived usefulness.

Model	Variable	*b*	*SE*	*t*	*p*	pr^2
1	Target Language					
1	Spanish vs French	0.05	0.42	0.12	.91	< .01
1	Spanish vs Other	-0.03	0.42	-0.08	.94	< .01
1	Course Level	-0.22	0.31	-0.71	.49	.04
1	Avg Final Grade	0.18	0.44	0.41	.86	.01
2	Age	0.06	0.06	1.01	.33	.08
2	Gender	0.77	0.71	1.08	.30	.09
3	GPA	-0.17	0.29	-0.61	.55	.03
4	Motivation	-0.01	0.04	-0.28	.79	.01
5	Mean Perceived Usefulness	0.69	0.21	3.30	.01	.55

Note. Target language was represented as three dummy variables with Spanish serving as the reference group; course level with introductory serving as the reference group; and gender with males serving as the reference group.

Discussion

The current study aimed to shed light on how metacognitive strategies can be used in a diverse set of structured language classes at the university setting. Specifically, we sought information about which metacognitive strategies could be taught and which, if any, would be most beneficial for such a diverse sample. This study was limited by small sample sizes. We present our study as an example of a potential type of research design to study metacognition within the university classroom, along with trends that metacognitive strategies could be beneficial for university students.

Prediction of Final Grade based on Voluntary Participation

The first hypothesis showed a moderate effect size for the unique variance accounted for by participation in the interventions alone ($\Delta R^2 = .06$). These results appear to indicate that learning metacognitive strategies could be beneficial when taught with the standard language learning material. When students are exposed to information about how to learn effectively, their success in learning outcomes increases (Dignath et al., 2008; Kramarski & Mevarech, 2003; Pishghadam & Khajavy, 2013; Vrugt & Oort, 2008; Wolters, 2004).

The most daunting limitation for specifically analyzing final grade as the sole learning outcome is grade inflation. Educators and researchers continually express their concern that university grades are not a reflection of actual learning (Palfreyman, 2010); Rojstaczer and Healy (2012) discovered that the proportion of A grades given increased from 15% of all letter grades in 1960 to 43% of all letter grades in 2009. A similar distribution of grades was found across the 469 students enrolled in the courses analyzed in the current study who received a letter grade for their coursework (i.e., did not withdraw or receive an incomplete grade); 47.5% received a final grade of A or A-. Grade inflation may be especially true for language learning courses, where student retention is already a common concern (Palfreyman, 2010; Rojstaczer & Healy, 2012). Future research should consider diverse learning outcomes, such as final course grade, standardized test scores, perceived accuracy, and perceived fluency, especially considering multiple outcomes simultaneously.

When deciding to analyze the data using regression and comparing to a randomly selected nonparticipant control group, demand characteristics became a potentially confounding variable. Some of the effects found in the current study may be due said characteristics. It is imperative that future research obtain a sample size large enough to have the necessary statistical power, and, potentially, the ability to perform multilevel models. Additionally, conducting studies using research designs which counterbalance or eliminate demand characteristics (such as longitudinal designs performed within the classroom as opposed to voluntary extra-curricular interventions) could improve the reliability and validity of the study.

Moreover, the current pilot study cannot guarantee that students exposed to interventions truthfully understood material presented during this time. Two to three manipulation check questions were presented for each intervention and time spent on the page was recorded for each participant, but correctly answering a short amount of multiple-choice questions does not provide sufficient evidence to claim true comprehension. Thus, some participants may have participated in the intervention, but not retained the information presented during the intervention. Future research should consider other ways to ensure comprehension, such as longer manipulation checks or time set aside specifically to practice metacognitive strategies.

Prediction of Persistence based on Voluntary Participation

The second hypothesis was not significant. Insufficient power due to small sample size may have influenced these results, as the effect size was moderate (Nagelkerke's Pseudo $R^2 = .20$). It was estimated that a sample of 153 would be best for the hierarchical binary log regression with $\alpha = .05$ and power = .80 (Faul et al., 2007); however, the achieved sample size was approximately half at $N = 86$. Despite lack of significance and power, more students who volunteered to participate in the metacognitive interventions continued their language learning endeavors compared to their matched nonparticipant peers.

Additionally, GPA was also unable to predict persistence. The cause for continued student enrollment is multi-faceted, and varies from individuals in different situations. While any course may be enjoyable and inspire a student to continue to enroll in similar courses, many students do not have the time or funds, such as one participant who expressed their reason for quitting their TL as, "It just won't fit in my schedule". Reasons for lack of persistence can vary outside the realm of lack of success.

Statistically, another limitation of this analysis is the difference in persistence group sizes. The number of students who did persist ($N = 44$) was almost double the number of students who did not persist ($N = 26$). While persistence is desirable, statistically, predicting correct group membership was much more likely for the larger group

(persistence; 81.8% correct predictions) than the smaller group (non-persistence; 48.2% correct predictions). The reasons for non-persistence vary between individuals and contexts, and the raw number of those not persisting was small, which created statistical barriers for making correct predictions. Future research should consider the necessary sample size to achieve not only sufficient power, but also large enough persistence group sizes to make accurate predictions for both groups. However, we theorize that the group of students choosing to not persist in language learning may continue to be smaller than the group of students persisting, which will further the statistical limit on predictions. Further, university characteristics, such as requirements for enrollment and major should be considered when examining persistence. The university where this research was performed required students to complete 12 hours in a single language beyond the first four introductory courses (i.e., 101, 102, 201, 202) for the Bachelor of Arts. To enroll in these courses, students must complete through 202, therefore, persistence is an important factor for degree completion.

Conclusion

Generally, providing students with the tools to understand how they learn most effectively can provide benefit for learning outcomes in most domains of learning (Cross & Paris, 1988; Dignath et al., 2008; Faul et al., 2007; Haller et al., 1988; Kramarski & Mevarech, 2003; Raoofi et al., 2013; Wolters, 2004). While sample sizes in the current study were smaller than recommended, many analyses still found moderate effect sizes. Thus, a trend towards efficacy in exposure to metacognitive interventions is evident and should be explored.

Further, while Remind 101 can provide quick information to an audience (such as reminders about exams or permission slips), it may not be the best tool for recruitment in other contexts (such as voluntary participation in time series research). Even when receiving three texts per week, participants were likely to ignore the messages. Future research should consider other methods of communication and/or stronger incentives.

Similarly, the attrition rate between time points may have been reduced if the interventions were woven into the current course material. That is, if educators could incorporate metacognitive skills within the language learning classroom, students would be more likely to be exposed to, and effectively learn, the information compared to strictly voluntary exposure. Future research should consider applying the current methods as an integral part of the language classroom. Likewise, providing "check your learning" time points wherein participants can actively see the strategies at work and prove that attempts are made to use said strategies could provide more efficacy for the research study.

While statistical analyses were unable to detect many statistically significant values in the current study, the moderate effect sizes allude that students may have much to gain by practicing metacognitive strategies in tandem with language skills. These strategies could improve language learning and allow classroom instruction to be most effective for all students, despite differences such as general aptitude, linguistic emotions, and learning preferences. Future research should focus on tackling the shortcomings of this pilot study, namely increasing sample size, reducing the effects of demand characteristics, and integrating the research design within the classrooms that would allow researchers to examine language performance in ways other than final grades. Educators should work to provide tools which educate students on metacognitive awareness and capitalize on the benefits of metacognition within the classroom, as providing metacognitive instruction has already been found to greatly improve learning outcomes in other academic areas (Cross & Paris, 1988; Dignath et al., 2008; Haller et al., 1988; Kramarski & Mevarech, 2003; Raoofi et al., 2013; Wolters, 2004), and improved learning outcomes have been found to increase rates of persistence (i.e., enrollment; Murtaugh, Burns, & Schuster, 1999). If educators can integrate metacognitive strategies in their classroom, students may be able to perform better, retain more language information, and continue enrolling in the TL.

References

Atkinson, R. C. (1975). Mnemotechnics in second-language learning. *The American Psychologist, 30*(8), 821–828. https://doi.org/10.1037/h0077029

August, D., Calderon, M., & Carlo, M. (2002). *Transfer of skills from Spanish to English: A study of young learners.* Center for Applied Linguistics. Retrieved from http://www.cal.org/acquiringliteracy/pdfs/skills-transfer.pdf.

Avila, E., & Sadoski, M. (1996). Exploring new applications of the keyword method to acquire English vocabulary. *Language Learning, 46*(3), 379–395. https://doi.org/10.1111/j.1467-1770.1996.tb01241.x

Brown, A. L. (1978). Knowing when, where, and how to remember: A problem of metacognition. *Advances in Instructional Psychology, 1,* 77–165.

Chenoweth, N. A., & Hayes, J. R. (2001). Fluency in writing: Generating text in L1 and L2. *Written Communication, 18*(1), 80–98. https://doi.org/10.1177/0741088301018001004

Coady, J., Magoto, J., Hubbard, P., Graney, J., & Mokhtari, K. (1993). High frequency vocabulary and reading proficiency in ESL readers. In T. Huckin, M. Haynes, & J. Coady (Eds.), S*econd language reading and vocabulary learning* (pp. 217–226). Norwood, NJ: Able.

Cross, D. R., & Paris, S. G. (1988). Developmental and instructional analyses of children's metacognition and reading comprehension. *Journal of Educational Psychology, 80*(2), 131–142. https://doi.org/10.1037/0022-0663.80.2.131

Dignath, C., Buettner, G., & Langfeldt, H.-P. (2008). How can primary school students learn self-regulated learning strategies most effectively? *Educational Research Review, 3*(2), 101–129. https://doi.org/10.1016/j.edurev.2008.02.003

Faul, F., Erdfelder, E., Lang, A.-G., & Buchner, A. (2007). G*Power 3: A flexible statistical power analysis program for the social, behavioral, and biomedical sciences. *Behavior Research Methods, 39*(2), 175–191. https://doi.org/10.3758/BF03193146

Flavell, J. H. (1979). Metacognition and cognitive monitoring: A new area of cognitive-developmental inquiry. *The American Psychologist, 34*(10), 906–911. https://doi.org/10.1037/0003-066X.34.10.906

Fraser, C. A. (1999). Lexical proessing strategy use and vocabulary learning through reading. *Studies in Second Language Acquisition, 21*(2), 225–241. https://doi.org/10.1017/S0272263199002041

Fukkink, R. G., Blok, H., & de Glopper, K. (2001). Deriving word meaning from written context: A multicomponential skill. *Language Learning, 51*(3), 477–496. https://doi.org/10.1111/0023-8333.00162

Gardner, R. C. (2004). Attitude/motivation test battery: International AMTB research project. The University of Western Ontario. Retrieved from http://publish.uwo.ca/~gardner/docs/englishamtb.pdf.

Gardner, R. C. (2007). Motivation and second language acquisition. *Porta Linguarum, 8,* 9–20. Retrieved from http://hdl.handle.net/10481/31616.

Gilmore, A. (2007). Authentic materials and authenticity in foreign language learning. *Language Teaching, 40*(2), 97–118. https://doi.org/10.1017/S0261444807004144

Goh, C. (2008). Metacognitive instruction for second language listening development: Theory, practice and research implications. *RELC Journal, 39*(2), 188–213. https://doi.org/10.1177/0033688208092184

Haller, E. P., Child, D. A., & Walberg, H. J. (1988). Can comprehension be taught? A quantitative synthesis of "metacognitive" studies. *Educational Researcher, 17*(9), 5–8. https://doi.org/10.2307/1175040

Hirsh, D., & Nation, P. (1992). What vocabulary size is needed to read unsimplified texts for pleasure? *Reading in a Foreign Language, 8,* 689–696.

Hodara, M., & Cox, M. (2016). *Developmental education and college readiness at the University of Alaska* (No. REL 2016–123). Retrieved from U.S. Department of Education, Institute of Education Sciences, National Center for Education Evaluation and Regional Assistance, Regional Educational Laboratory Northwest website: http://eric.ed.gov/?id=ED565798

Hodara, M., & Lewis, K. (2017). *How well does high school grade point average predict college performance by student urbanicity and timing of college entry?* Retrieved from Regional Educational Laboratory Program website: https://ies.ed.gov/ncee/edlabs/projects/project.asp?projectID=4546

Hong-Nam, K., & Leavell, A. G. (2006). Language learning strategy use of ESL students in an intensive English learning context. *System, 34*(3), 399–415. https://doi.org/10.1016/j.system.2006.02.002

Hsiao, T., & Oxford, R. L. (2002). Comparing theories of language learning strategies: A confirmatory factor analysis. *The Modern Language Journal, 86*(3), 368–383. https://doi.org/10.1111/1540-4781.00155

Huckin, T., & Bloch, J. (1993). Strategies for inferring word meaning in context. In T. Huckin, M. Haynes, & J. Coady (Eds.), *Second language reading and vocabulary learning* (pp. 153–178). Norwood, NJ: Able.

Karpicke, J. D. (2009). Metacognitive control and strategy selection: deciding to practice retrieval during learning. Journal of Experimental Psychology. General, 138(4), 469–486. https://doi.org/10.1037/a0017341

Kramarski, B., & Mevarech, Z. R. (2003). Enhancing mathematical reasoning in the classroom: The effects of cooperative learning and metacognitive training. *American Educational Research Journal, 40*(1), 281–310. https://doi.org/10.3102/00028312040001281

Kulhavy, R. W., & Swenson, I. (1975). Imagery instructions and the comprehension of text. The British Journal of *Educational Psychology, 45*(1), 47–51. https://doi.org/10.1111/j.2044-8279.1975.tb02294.x

Laufer, B. (1997). What's in a word that makes it hard or easy: Some interalexical factors that affect the learning of words. In N. Schmitt & M. McCarthy (Eds.), *Vocabulary: Description, acquisition and pedagogy* (pp. 140–180). Cambridge: Cambridge University Press.

Moir, J., & Nation, I. S. (2002). Learners' use of strategies for effective vocabulary learning. *Prospect, 17,* 15–35.

Moore, J. C., & Surber, J. R. (1992). Effects of context and keyword methods on second language vocabulary acquisition. *Contemporary Educational Psychology, 17*(3), 286–292. https://doi.org/10.1016/0361-476x(92)90067-9

Morrison, L. (1996). Talking about words: A study of French as a second language learners' lexical inferencing procedures. *Canadian Modern Language Review, 53*(1), 41–75. https://doi.org/10.3138/cmlr.53.1.41

Murtaugh, P. A., Burns, L. D., & Schuster, J. (1999). Predicting the retention of university students. *Research in Higher Education, 40*(3), 355–371. https://doi.org/10.1023/A:1018755201899

Nassaji, H. (2003). L2 vocabulary learning from context: Strategies, knowledge sources, and their relationship with success in L2 lexical inferencing. *TESOL Quarterly, 37*(4), 645–670. https://doi.org/10.2307/3588216

Nation, I. S. P., & Webb, S. A. (2011). *Researching and analyzing vocabulary.* Heinle & Heinle Pub.

Nyikos, M., & Oxford, R. (1993). A factor analytic study of language-learning strategy use: Interpretations from information-processing theory and social psychology. *The Modern Language Journal, 77*(1), 11–22. https://doi.org/10.1111/j.1540-4781.1993.tb01940.x

O'Malley, J. M., & Chamot, A. U. (1990). *Learning strategies in second language acquisition*. Cambridge University Press. https://doi.org/10.1017/CBO9781139524490

O'Malley, J. M., Michael O'Malley, J., Chamot, A. U., Stewner-Manzanares, G., Russo, R. P., Küpper, L., & Kupper, L. (1985). Learning strategy applications with students of English as a second language. *TESOL Quarterly, 19*(3), 557–584. https://doi.org/10.2307/3586278

Oxford, R. L. (1990). *Language learning strategies: What every teacher should know.* Newbury House Publishers.

Palfreyman, D. (2010). Excellence without a soul: How a great university forgot education. *Perspectives: Policy and Practice in Higher Education, 14*(1), 29–31. https://doi.org/10.1080/13603100802002701

Pishghadam, R., & Khajavy, G. H. (2013). Intelligence and metacognition as predictors of foreign language achievement: A structural equation modeling approach. *Learning and Individual Differences, 24,* 176–181. https://doi.org/10.1016/j.lindif.2012.12.004

Pressley, M., Levin, J. R., & Delaney, H. D. (1982). The mnemonic keyword method. *Review of Educational Research, 52*(1), 61–91. https://doi.org/10.3102/00346543052001061

Raoofi, S., Chan, S. H., Mukundan, J., & Rashid, S. M. (2013). Metacognition and second/foreign language learning. *English Language Teaching, 7*(1), 36–49. https://doi.org/10.5539/elt.v7n1p36

Reason, R. D. (2003). Student variables that predict retention: Recent research and new developments. *NASPA Journal, 40*(4), 172–191. https://doi.org/10.2202/0027-6014.1286

Robertson, K. (2010). Improving writing skills: ELLs and the joy of writing. Retrieved from Colorin Colorado website: https://www.colorincolorado.org/article/improving-writing-skills-ells-and-joy-writing

Robson, G., & Midorikawa, H. (2001). How reliable and valid is the Japanese version of the Strategy Inventory for Language Learning (SILL)? *JALT Journal, 23*, 202–226.

Rojstaczer, S., & Healy, C. (2012). Where A is ordinary: The evolution of American college and university grading. *Teachers College Record, 114,* 1–23.

Schatz, E. K., & Baldwin, R. S. (1986). Context clues are unreliable predictors of word meanings. *Reading Research Quarterly, 21*(4), 439–453. https://doi.org/10.2307/747615

Schoonen, R., van Gelderen, A., Glopper, K. de, Hulstijn, J., Simis, A., Snellings, P., & Stevenson, M. (2003). First language and second language writing: The role of linguistic knowledge, speed of processing, and metacognitive knowledge. *Language Learning, 53*(1), 165–202. https://doi.org/10.1111/1467-9922.00213

Schraw, G., Crippen, K. J., & Hartley, K. (2006). Promoting self-regulation in science education: Metacognition as part of a broader perspective on learning. *Research in Science Education, 36*(1-2), 111–139. https://doi.org/10.1007/s11165-005-3917-8

Vandergrift, L. (2004). Listening to learn or learning to listen? *Annual Review of Applied Linguistics, 24,* 3–25. https://doi.org/10.1017/S0267190504000017

Veenman, M. V. J., Van Hout-Wolters, B. H. A. M., & Afflerbach, P. (2006). Metacognition and learning: conceptual and methodological considerations. *Metacognition and Learning, 1*(1), 3–14. https://doi.org/10.1007/s11409-006-6893-0

Vrugt, A., & Oort, F. J. (2008). Metacognition, achievement goals, study strategies and academic achievement: pathways to achievement. *Metacognition and Learning, 3*(2), 123–146. https://doi.org/10.1007/s11409-008-9022-4

Wang, A. Y., Thomas, M. H., & Ouellette, J. A. (1992). Keyword mnemonic and retention of second-language vocabulary words. *Journal of Educational Psychology, 84*(4), 520–528. https://doi.org/10.1037/0022-0663.84.4.520

Wimolmas, R. (2013). *A survey study of motivation in English language learning of first year undergraduate students at Sirindhorn International Institute of Technology.* Language Institute, Thammasat University.

Wolters, C. A. (2004). Advancing achievement goal theory: Using goal structures and goal orientations to predict students' motivation, cognition, and achievement. *Journal of Educational Psychology, 96*(2), 236–250. https://doi.org/10.1037/0022-0663.96.2.236