An Investigation Into Second Language Aptitude for Advanced Chinese Language Learning

PAULA WINKE
Michigan State University
Second Language Studies Program
Department of Linguistics and Languages
B252 Wells Hall, 619 Red Cedar Road
East Lansing, MI 48824
Email: winke@msu.edu

In this study I examine the construct of aptitude in learning Chinese as a second language (L2) to an advanced level. I test 2 hypotheses: first, that L2 aptitude comprises 4 components—working memory, rote memory, grammatical sensitivity, and phonemic coding ability—and second, that L2 aptitude affects learning both directly and indirectly (mediated by strategy use and motivation). Native speakers of English (n = 96) studying advanced Chinese took the Modern Language Aptitude Test and a phonological working memory test and responded to motivation and strategy use questionnaires. Using end-of-course listening, reading, and speaking proficiency test results as measures of Chinese learning, I constructed a structural equation model to test the hypotheses. The model fit the observed data. Of the 4 components foreseen to comprise L2 aptitude, rote memory contributed the most and working memory the least. Aptitude, strategy use, and motivation had about the same impact on learning but varied in how well they predicted the individual skills of listening, reading, and speaking. The results shed light on L2 aptitude in the particular context of an advanced L2 Chinese course.

COGNITIVE PROCESSES AND AFFECTIVE variables shape how individuals acquire a foreign or second language (L2) and predict how well they are likely to learn one (Beckner et al., 2009; Ellis, 2004; Robinson, 2002c; Skehan, 1989). This study investigates the plausibility, via structural equation modeling (SEM), of a model of language learning that includes cognitive (rote memory, phonemic coding ability, grammatical sensitivity, and phonological working memory), cognitively oriented (strategy use), and affective (motivation) variables as learning predictors. It also examines how the factors affect each other within the model.

First, I review research that addresses what I suggest is a narrow definition of L2 aptitude as a set of purely cognitive constructs. Second, I argue that a broader concept of L2 aptitude should be accepted, one that recognizes the effects of mediating, cognitively oriented, and/or affective variables—most vitally, strategy use and motivation. Third, because L2 aptitude may be best understood in terms of the context of the language learning situation (Robinson, 2007), I detail the particular conditions of this study’s L2 learning context: native-English-speaking adults learning Chinese in an intensive (6 to 8 hours a day in class), 63-week course intended to bring them to an advanced level of proficiency in Chinese. Finally, I describe the use of SEM to investigate L2 learning aptitude in this context, which has been rarely used in L2 aptitude research.

THE COGNITIVE CONSTRUCTS OF SECOND LANGUAGE APITUDE

Acquiring proficiency in an L2 in an instructed setting is considered challenging for adults (Doughty, 2004; Ellis, 2004, 2005). Understanding
why has engrossed L2 acquisition researchers for decades. Since the 1950s, researchers have posited that L2 aptitude is a construct separate from general intelligence and predicts adult, classroom-based, L2 learning success (Carroll, 1962, 1981, 1990; Corno et al., 2002; Dörnyei, 2005b; Robinson, 2005a; Skehan, 2002). The variables that were first recognized as part of L2 aptitude were identified prior to and/or independent of modern second language acquisition (SLA) theory. Carroll (1958), tasked with creating an L2 aptitude test for the U.S. government, administered 34 cognitive tests to military personnel taking intensive, 1-week-trial, beginner-level, Mandarin Chinese courses that focused on speaking skills. By using exploratory factor analysis, Carroll identified five tests that were both practical to administer and highly predictive of L2 learning success. These tests now comprise the Modern Language Aptitude Battery, or MLAT, Form A (Carroll & Sapon, 1959b). Three of the five tests’ underlying constructs identified by Carroll (1958, 1962) will be explored here: rote memory for learning (the ability to store verbal information in memory and recall it later); phonetic coding ability (the ability to analyze incoming sounds so that they can be retained); and grammatical sensitivity (the ability to recognize, in whatever manner, the function of a word in a sentence). (For a more detailed description of these constructs, see Nagata, Aline, & Ellis, 1999, p. 135, and Skehan, 1998, p. 190.) Carroll followed this research with several MLAT validation studies (Carroll, 1962, 1963, 1966) and concluded that under intensive learning conditions with homogeneous groups of learners, the MLAT correlates fairly well with L2 achievement, with most correlation coefficients between .40 and .65 (cf. Carpenter, 2009; Skehan, 1998).

Over the decades, the predictive validity of the MLAT has prevailed. The MLAT is still used by the U.S. Foreign Service and the Canadian Foreign Service in selecting candidates for L2 study. The MLAT is also used in research that attempts to define the cognitive traits that explain differences in adult L2 acquisition (e.g., research on critical period effects, DeKeyser, 2000; Harley & Hart, 1997, 2002; Ross, Yoshinaga, & Sasaki, 2002). Other aptitude tests, based on the MLAT or shown to be valid through correlations with the MLAT, have been developed (e.g., the Cognitive Ability for Novelty in Acquisition of Language [Foreign] Test or CANAL–FT, Grigorenko, Sternberg, & Ehrman, 2000; the Defense Language Aptitude Battery or DLAB, Peterson & Al–Haik, 1976; the Pimsleur Language Aptitude Battery or PLAB; and other measures). But differing tests of L2 aptitude do not fundamentally alter our understanding of the underlying cognitive constructs of L2 aptitude (Skehan, 1998). Moreover, they do not extend the construct of L2 aptitude beyond what the tests measure (Dörnyei, 2005b; Neufeld, 1979; Sáfár & Kormos, 2008).

It may profit the field to take aptitude research to another level and investigate aptitude for dynamic complexity. For example, Skehan (2002) proposed that aptitude abilities are dynamic and change over time. Robinson (2005a, 2007) suggested that different aptitude abilities are needed at different stages of learning. Recent theorizing in SLA has proposed that learning is fundamentally a slave to the environment, with certain contexts nourishing some mental processes and other contexts nourishing other mental processes (Beckner et al., 2009; Dörnyei, 2009a; Larsen–Freeman & Cameron, 2008). It is not, therefore, an understatement to claim that aptitude as a standalone cognitive trait is problematic. One place to start is to investigate aptitude constructs individually and within a larger picture of cognitive, cognitively oriented, and affective variables and situated within a particular language learning context. We can at least then begin to understand if aptitude is dynamic and relates to the learning context, as has been suggested (Neufeld, 1979; Snow, 1987; Sternberg, 2002; Sternberg & Grigorenko, 2002).

The MLAT

Many researchers have challenged the view that L2 aptitude is something measured by the MLAT (Dörnyei, 2005b; Robinson, 2002b, 2007; Sáfár & Kormos, 2008; Skehan, 2002). One criticism is that the constructs underlying the MLAT do not represent a complete definition of L2 aptitude. This has been confirmed by research that has shown that the MLAT does not strongly predict language learning when instruction is less intense (Carroll, 1962, 1990) or more communicative in nature (Robinson, 2007; Sáfár & Kormos, 2008). Furthermore, in the original validation studies by Carroll, learners only received lower level instruction. Thus, even though there have been decades of studies on L2 aptitude, we still do not know what cognitive variables are important for learning in moderately intense courses, in communicative, task-based language programs, or for achieving advanced L2 skills.

This does not mean that researchers should abandon the MLAT but rather that the
constellation of factors that contribute to successful language acquisition needs to be expanded beyond the constructs represented by the MLAT and more firmly situated within the context of learning. Furthermore, aggregating L2 aptitude test subscores into a single composite score (which indirectly implies what L2 aptitude is—a single factor) needs to be revisited. Skehan (1998, 2002) called for researchers to view aptitude as truly differentiated or multifaceted, meaning that one could have high ability in one aptitude construct and low ability in others, resulting in individually unique L2 aptitude profiles. This is different from Carroll’s view that there are across-the-board haves and have-nots in L2 aptitude or that L2 aptitude is “a special, inherent talent that not all individuals possess” (CASL, 2009, p. 1). Robinson (2007) proposed the Ability Differentiation Hypothesis, which claims that some L2 learners have clearly differentiated cognitive skills and abilities, while others do not, and that these different talent schemes correspond with different aptitude complexes, which must be matched to instructional conditions to maximize L2 learning potential. According to Skehan and Robinson, within the model, the individual components may be correlated for some learners, but not for others.

**Working Memory**

In the 1990s, building on research in cognitive psychology that identified working memory as a cognitive trait separate from long- and short-term memory systems and important for learning (Baddeley, 1986, 1992, 2007; Cowan, 1995, 2005; Daneman & Carpenter, 1980; Hitch & Baddeley, 1976), applied linguists investigated the relationships between working memory and L2 acquisition and ascertained that working memory is an essential component for L2 learning (Gathercole & Baddeley, 1993; Harrington & Sawyer, 1990, 1992; Miyake & Friedman, 1998; Miyake, Friedman, & Osaka, 1998; Service, 1992). Concomitantly, researchers suggested that working memory is most likely an additional cognitive construct of L2 aptitude (Daneman & Merikle, 1996; Harrington & Sawyer, 1992; McLaughlin, 1995; Miyake & Friedman, 1998; Robinson, 1995, 2002b) and may be the “key to elaborating the concept of language aptitude itself” (Sawyer & Ranta, 2001, p. 340).

Generally defined, working memory underlies the ability to process linguistic input and store information from that input for later retrieval. Specifically, working memory has been defined as the “set of processes that hold a limited amount of information in a readily accessible state for use in an active task” (Cowan, 2005, p. 39). In other words, working memory is a cover term for multiple processes, including short-term memory, the real-time manipulation of linguistic material through effortful processing (Cowan, 2005) and storing information in long-term memory. These are dissociable abilities, each with limited capacities, but that must work together for learning (Baddeley, 2007; Miyake & Friedman, 1998).

Empirical investigation into working memory’s role in relation to previously identified L2 aptitude constructs is of current interest in the field for both practical and theoretical reasons. Several government-sponsored research projects support such research with the goal to increase the government’s ability to identify military and foreign service candidates for language training (see CASL, 2009; Kenyon & MacGregor, 2004). Theoretically, researchers are interested because they need to understand L2 aptitude as a construct and how it fits within other theories about learning (Corno et al., 2002). Thus far, research suggests that working memory is an essential part of L2 aptitude (Robinson, 2002b; Sáfár & Kormos, 2008), but evidence is limited. Some research suggests that it may only significantly differentiate learning at lower levels of instruction (Hummel, 2009). Other research that focuses on the short-term memory component of working memory suggests that limited short-term memory capacity may impede learning for those with limited capacities while failing to differentiate other learners (Carpenter, 2009). Such findings question the entire concept of L2 aptitude: That is, is there really a single latent trait that can reliably predict advanced-level language skills or L2 skills acquired through nonintense or more naturalistic conditions? Researchers may have pinned their hopes on working memory as a promising additional component of L2 aptitude that reliably predicts advanced proficiency, but initial research is inconclusive. More studies are needed.

**SECOND LANGUAGE APTITUDE AND NONCOGNITIVE AND AFFECTIVE VARIABLES**

It is simplistic to think that language learning depends solely on cognitive traits. Researchers have long attributed language learning success to a number of noncognitive factors, such as high motivation for learning the particular language at hand (Dörnyei, 1990; Gardner, 1990, 2001; Gardner & Lambert, 1959; Gardner, Tremblay,
Motivation

Motivation affects L2 acquisition in multiple ways. It is a necessary precondition for L2 acquisition (Csizér & Dörnyei, 2005; Dörnyei, Csizér, & Németh, 2006; Edmondson, 2004; Noels et al., 2000), and it must be healthily sustained over time for acquisition to continue (Dörnyei, 2005b, 2009b; Hiromori, 2009). For over three decades, motivation has been shown to be a predictor of L2 learning success (Csizér & Dörnyei, 2005; Gardner, 1985; Gardner & MacIntyre, 1991; Noels, Clément, & Pelletier, 1999; Schmidt & Watanabe, 2001; Tremblay & Gardner, 1995). These studies and others have conceptualized motivation differently, but most view it from a social–psychological perspective.

In other words, a highly motivated L2 learner will (a) want to integrate with speakers of the language (integrativeness is the desire to become a passable member of the community of speakers of the L2) (Gardner, 1985, 2001; Gardner et al., 1997), (b) have a very good attitude about learning the L2 (Dörnyei, 1990, 2003; Yashima, Zenuk-Nishide, & Shimizu, 2004), and (c) be eager to communicate in the target language (Dörnyei & Kormos, 2000; MacIntyre et al., 2001; Yashima, 2002; Yashima et al., 2004). A highly motivated language learner is further characterized by strong and clear incentives for learning the language, such as for future employment, a pay raise, travel opportunities, or to communicate with family members (these are classified as instrumental and/or intrinsic reasons; for definitions see Gardner, 1985; Gardner, 2001; Gardner & MacIntyre, 1991; Gardner et al., 1997; Tremblay & Gardner, 1995). The individual will also have higher confidence in his or her abilities in using the L2 (Clément, Dörneyi, & Noels, 1994; Dörnyei, 2003, 2005a; Ushida, 2005)—and correspondingly lower anxiety in using the language (MacIntyre et al., 2002; Matsuda & Gobel, 2004). The overall consensus is that high motivation may make up for deficiencies in cognitive abilities (Dörnyei, 1990; Pimsleur, 1966; Schmidt, 1991; Sternberg, 2002), which suggests that the effects of cognitive abilities on L2 learning are mediated by motivation.

Strategy Use

Similarly, strategy use has been shown to affect L2 acquisition, but it is unclear how. Strategies are the “steps or actions taken by the learners to improve the development of their language skills” (Oxford & Cohen, 1992, p. 1) or “the L2 learner’s toolkit for active, conscious, purposeful, and attentive learning” (Hsiao & Oxford, 2002, p. 372). More recently, strategies have been viewed as cognitively oriented—as summarized by Macaro (2006), “strategies are the raw material of conscious cognitive processing, and their effectiveness or noneffectiveness derives from the way they are used and combined in tasks and processes” (p. 325). Studies that have measured strategy use through questionnaires have shown that learners who use many different kinds of strategies, and use them often, have more success in instructed L2 acquisition settings (Cohen, 1998; Ehrman & Oxford, 1990; Green & Oxford, 1995; Griffiths, 2003; O’Malley & Chamot, 1990; Oxford, 1990a, 1994). (See Oxford and Burry–Stock, 1995, for a full review.) Other research has found that less successful learners use various, random strategies, while more successful learners are more systematic about their strategy usage and use specific ones for specific tasks (Ehrman & Oxford, 1995). More recent research has suggested that strategy use changes over time depending on L2 proficiency and the learning context (Macaro, 2006), with
advanced L2 learners using fewer strategies overall than intermediate learners (Hong–Nam & Leavell, 2006) because at the advanced levels of proficiency, learners’ processes for L2 acquisition are more automatized, resulting in a smaller range of strategies needed for acquisition (see Oxford, 2011, for a comprehensive review of the research).

Most interestingly, research has shown that L2 learning strategies and motivation are related (Schmidt & Watanabe, 2001). More strategy use equals more or better learning, and more motivated learners use more learning strategies more often (Gardner et al., 1997; MacIntyre & Noels, 1996). Thus, a full model of L2 learning that considers the impact of L2 aptitude on learning should also investigate the combined mediating role of motivation and strategy use. As far as I know, no study thus far has done so.

THE SECOND LANGUAGE LEARNING CONTEXT

One of the problems with L2 aptitude research is that it often assumes that L2 aptitude exists in a vacuum and is independent from the L2 learning context. This may not be true, especially if we assume motivation and strategy use mediate aptitude’s effects on language learning. One might be more motivated to learn a certain L2 over another (Sternberg, 2002); likewise, one might be more motivated to learn an L2 in a certain classroom situation over another. This corresponds with what is known as the resultative or spiraling effect of motivation: Learners who do well in a certain learning situation become even more motivated to learn, while those who do poorly get discouraged and try less hard (Hermann, 1980). Classroom-based variables such as materials, instructional techniques, teachers, and peers might affect motivation and strategy use, which may positively impact learning, min- imizing or distorting the effects of L2 aptitude. As stated by Corno et al. (2002), “to understand the effects of person characteristics on performance, one must specify the performance situation” (p. 216). Thus, researchers must investigate L2 aptitude in terms of the context of the L2 learning situation (Robinson, 2007).

This current study is unique in that all learners stemmed from the same learning context. The learners began with no prior Chinese language instruction. All received the same instruction throughout a 63-week course, which was intended to develop in each language learner advanced-level proficiency (also known as General Professional Proficiency—a 3 on the Interagency Language Roundtable scale; see http://www.govitr.org for a description) in the skills of listening, reading, and speaking so that the learners could go on to be interpreters, analysts, and interrogators for the military branches for which they worked. All students received instruction from the same teachers, used the same materials, and followed the same curriculum.

Equal instructional conditions across participants is important because prior research has found that L2 aptitude is sensitive to exposure conditions, such as the amount of implicit versus explicit instruction (Robinson, 1997, 2002c, 2005b). Prior studies on aptitude may have had significant sampling errors in relation to instructional variance. For example, Hummel (2009) explored the relationships among the cognitive constructs of L2 aptitude (phonetic coding ability, grammatical sensitivity, rote memory, and phonological working memory) and L2 proficiency. She found that the cognitive constructs of L2 aptitude combined predicted 29% of the variance in L2 (English) proficiency. The 77 participants in her study were recruited from a first-year Teaching English as a Second Language degree program. L2 proficiency was measured directly after entrance (within their first month of study) in the program; the tests of L2 aptitude, including working memory, were administered at some point after the proficiency test (Hummel did not indicate when). Although Hummel described the participants as being a homogeneous group—all were native French speakers with at least 7 years of English study—the quality and methods of their prior English language instruction were not controlled. One could imagine a situation in which those who attended better language programs received better instruction on strategy use (which has been shown to improve L2 learning—see Plonsky, 2011) and were more motivated, which might have contributed to the relationships observed. Furthermore, if the participants attended differentially performing schools, with school access determined through a college entrance exam (another cognitive test or an intelligence test), any observed correlations between L2 aptitude and proficiency could be interpreted as noncausal: Other factors (i.e., general intelligence, instructional differences) may underlie both high aptitude and high performance. This is not a fatal research design flaw, but it allows one to question the reliability of such a study. It would be better to make certain that all learners received the same instruction (and the same amount of instruction), which
would help ensure that only the variables present in the hypothesized model of L2 learning were responsible for the study’s outcomes.

**USING STRUCTURAL EQUATION MODELING TO INVESTIGATE SECOND LANGUAGE APTITUDE**

The relationships among L2 aptitude constructs have been explored through correlation analyses (e.g., Wesche, Edwards, & Wells, 1982), factor analyses (e.g., Carroll, 1958, 1962, 1963, 1966, 1993), and simple and multiple regression (e.g., Carpenter, 2009; Hummel, 2009). However, these are essentially descriptive techniques, making complete hypothesis testing difficult (Byrne, 2009). For example, with confirmatory factor analysis (CFA), even though the researcher postulates how the underlying latent variables relate and then tests this structure statistically, CFA focuses on whether and the extent to which the observed variables are linked to the underlying latent traits. The strengths of the regression paths (the factor loadings) from the latent variables to the observed variables (the items) are the focus of CFA. CFA does not consider direct effects among the factors, and thus CFA misses essential elements—a full latent variable model (such as SEM) has. SEM combines a CFA measurement model (SEM runs a CFA as part of its analysis) and a structural model. In other words, a full SEM model allows researchers to estimate both the links between the latent variables and their observed measures (the measurement portion of the model) and the direct effects among the variables (the structural portion of the model).

With SEM researchers can investigate the plausibility of a full latent variable model, which is defined as a single proposed set of relationships among one or more independent variables and one or more dependent variables (Byrne, 2009). The relationships within the proposed model can be discussed in terms of their causality, but the cause–effect relationships must be backed up by relevant theory—that is, the proposed set of relationships in a model of L2 aptitude and L2 learning must stem from hypotheses and theories on L2 aptitude and L2 learning. As explained by Lewis and Vladeanu (2006):

> While there are similarities between structural equation modeling and multiple regression (e.g., they are based on analysis of intercorrelations) there are fundamental differences. Multiple regression provides simple definitive results about which independent variables produce significant effects on which dependent variables. In structural equation modeling, the analysis requires a theoretical model that can be tested against the available data. It can be determined whether this model is consistent with the data and also how good a fit it is. (p. 981)

A model that is found to be consistent with the data through SEM has nothing to say about whether a better model might be possible with the existing set of data or with other data. Thus, the results stemming from SEM are grounded within the context of the data set (in this case, within the data provided by adults learning Chinese in a communicative, task-based, 63-week course that focuses on the skills of listening, reading, and speaking). But, with SEM, competing models can be compared, a model can be tested over time, or a model can be tested across different data sets. This is exactly what L2 aptitude research needs, especially because theory now suggests that cognitive abilities may be represented differently by different individuals, may change over time, or may depend on the particular dynamics of the first language (L1)–L2 learning context. SEM also allows structural relations to be plotted in a picture, making it easy to visualize the theory being studied (Byrne, 2009). While other studies on L2 aptitude have employed SEM (i.e., Sasaki, 1993, 1996), this is the first L2 aptitude study using SEM that ensures participants received the same amount and type of instruction.

**HYPOTHEZIZED MODEL**

This study’s hypothesized model is presented in Figure 1. What I hypothesize is a model of L2 learning that includes both cognitive and affective traits. L2 aptitude is defined as comprising the cognitive traits of rote memory, phonetic coding ability, grammatical sensitivity, and working memory. L2 aptitude affects L2 learning both directly and indirectly—the indirect effects coming from mediation by the affective traits of strategy use and motivation.

**METHOD**

**Participants**

Ninety-six native English-speaking, adult learners of Chinese volunteered to be participants in this study. All were students in a 63-week, intensive Chinese program at the Defense Language Institute (DLI) in Monterey, California. All participants were U.S. military personnel. Sixty-six were male, 30 were female. Ages ranged from 19 to 36, with an average age of 25. As explained above, all received the same task-based, L2 instruction in
Chinese and received 6 to 8 hours of classroom-based instruction 5 days a week during the 63-week course.

Materials

Modern Language Aptitude Test (MLAT). The learners’ phonetic coding ability, grammatical sensitivity, and rote memory were assessed by sections one through five of the paper-and-pencil-based MLAT (Carroll & Sapon, 1959a). Based on Carroll’s (1962, 1990) and Skehan’s (1998, 2002) view of the underlying constructs of L2 aptitude assessed by the MLAT, rote memory was measured by sections one (number learning) and five (paired associates), phonetic coding ability by sections two (phonetic script) and three (spelling clues), and grammatical sensitivity by section four (words in sentences). The MLAT was obtained in 2004 through the Language Learning and Testing Foundation (http://lltf.net). For information concerning the five sections and their specific tasks, see Dörnyei (2005b) and the Language Learning and Testing Foundation’s (2012) website.

Phonological Working Memory Span Test. The participants’ working memory capacity was assessed through an online, verbal working memory span test adapted from Mackey et al. (2002),7 which was based on other span tests used in prior research (Daneman & Carpenter, 1980; Turner & Engle, 1989; Waters & Caplan, 1996). The test comprised 48 unrelated sentences, half being grammatically correct and half semantically plausible, which resulted in four sentence types (grammatical and plausible, grammatical and implausible, ungrammatical and plausible, ungrammatical and implausible). The 48 sentences were grouped into 12 sets: four sets each of three, four, and five sentences, groupings similar to those used by Daneman and Carpenter (1980) and Turner and Engle (1989). The test was self-paced; however, sentences within a set were controlled to play over the computer with 3-second intervals. While the participants listened to a set of sentences (the sentences were only presented aurally with no visual support), they clicked a mouse to indicate whether the sentence was grammatical and whether it was semantically plausible. After selecting “enter” to submit the answers for a set, the participants were prompted to type (on a new screen) the last word of each sentence. All sentence-final words were common (not abstract, because abstract words may be more difficult to recall; Turner & Engle, 1989), noncompound, concrete nouns of one to three syllables in length. No sentence-final word of a given set was semantically associated with another word in that set, and no words within a set rhymed.

Strategy Inventory for Language Learning (SILL). Strategy use was measured through the Strategy Inventory for Language Learning, or SILL (Oxford, 1990a), which was administered online for this study. The SILL comprises 80 five-point, Likert-scale questions pertaining to (a) social, (b) metacognitive, (c) memory, (d) compensation, (e) cognitive, and (f) affective strategies (Hsiao & Oxford, 2002; Oxford, 1990a). It has been used in numerous studies on language learning strategies (e.g., Carson & Longhini, 2002; Engelbar & Theuerkauf, 1999; Green &
Oxford, 1995; Griffiths, 2003; Hong-Nam & Leavella, 2006; Hwu, 2007; Nakatani, 2006; Oxford & Burry-Stock, 1995; Wharton, 2000) and measures the number and type of strategies applied by the learner at a specific point in the learning process (Tseng & Schmitt, 2008).

**Motivation Questionnaire.** The motivation questionnaire consisted of 38 five-point, Likert-scale questions adapted from Kormos and Dörnyei (2004). The questions, which tap into learners’ (a) integrativeness, (b) incentive values, (c) attitudes toward learning the L2, (d) linguistic self-confidence, (e) language use anxiety, (f) task attitudes, and (g) willingness to communicate (Dörnyei, 2002) have been used in multiple research studies on motivation for learning an L2 (Dörnyei, 2002; Dörnyei & Kormos, 2000; Kormos & Dörnyei, 2004; Weger–Guntihar, 2008). The questions appear in Appendix A.

**L2 Learning.** L2 learning was measured through the Defense Language Proficiency Tests (DLPTs) in listening, reading, and speaking. Robust descriptions of and sample items from these secure, government tests can be found at the Defense Language Institute Foreign Language Center (n.d.) Web site. These test scores can be considered gain scores because all study participants began the 63-week Chinese course with no prior instruction in Chinese.

**Procedure**

Data collection took place at the DLI in Monterey, California, during the learners’ 42nd week of instruction. All participants volunteered to participate in the research, which was conducted outside of normal class time. In computer labs on campus, participants took the following measures with 15- to 20-minute breaks in between: the MLAT, the working memory span test, the motivation questionnaire, and the SILL. Data collection, with breaks, lasted approximately 2 1/2 hours. After 63 weeks of instruction, the participants took the DLPTs in speaking, reading, and listening. The DLPT scores were forwarded to the author for use in the structural equation model.

**Scoring.** The MLAT was scored conventionally, with each item on the test scored as right or wrong, and a right answer worth 1 point. Rote memory was a composite of sections one and five. Phonetic coding ability was a composite of sections two and three. Grammatical sensitivity was measured by section four.

The phonological working memory test was scored according to partial-credit load (PCL) scoring procedures outlined in Conway et al. (2005). For any given set of sentences, if processing scores fell below 80%, the data from that set were discarded. If processing scores were above 80%, then all sentence-final words recalled correctly were awarded 1 point. Therefore, items with a higher memory load contributed more to the overall phonological working memory score. As summarized by Conway et al., “for load-weighted scoring procedures, PCL represents the sum of correctly recalled elements from all items, regardless of whether the items are perfectly recalled or not (also without respect to serial order within items)” (p. 775).

Items on the SILL were combined into a composite SILL score to indicate what normally is considered a latent variable. The same was done with the items on the motivation questionnaire. I did this to make the whole model easier for the computer program Analysis of Moment Structures (AMOS) 18 (structural equation modeling software) to identify. For each learner (and for each measure), items pertaining to the same category were averaged, and then all categories were averaged together so that each category would have equal weight in the learner’s final score. (The categories for each measure are described above in the materials section.) In both questionnaires, values of negative items were reversed before the aggregation.

**Analysis.** I applied SEM to evaluate the conjectured causal relations among the various variables investigated in the study. First, a full model with both measurement and structural components was designed in accordance with the theories in L2 aptitude and L2 learning reviewed earlier (see Figure 2). The measurement portion of the model hypothesizes cause indicators, rather than the more common effect indicators (Bollen, 1989, p. 65). Maximum likelihood estimated procedures were used to analyze the variance/covariance matrix of the observed variables using AMOS 18. To assess the overall model fit, I used chi-square and a pair of fit indices advised in the SEM literature (Byrne, 2009; Kline, 2011; McDonald & Ho, 2002): the comparative fit index (CFI) and the root-mean-square error of approximation (RMSEA). A nonsignificant chi-square and a CFI above .95 suggest model acceptance, and an RMSEA value below .05 indicates a good fit of the model to the data (Hu & Bentler, 1999).

**RESULTS**

**Preliminary Data Analysis**

The means, standard deviations, and reliabilities (Cronbach’s alpha) for the different measures are
in Table 1. Reliabilities for the different measures were mostly high or very high (Bachman & Palmer, 2010), ranging from .72 to .91. The sample correlation matrix is presented in Appendix B; the matrix allows readers to form an independent judgment of the relationships among the observed variables and will aid in the interpretation of the direct effects among the variables in the structural portion of the model. Rote memory was significantly correlated with both grammatical sensitivity ($r = .45, p = .00$) and phonological working memory ($r = .23, p = .04$). Strategy use significantly correlated with motivation ($r = .40, p = .00$). Reading was significantly related to both listening ($r = .59, p = .00$) and speaking ($r = .32, p = .00$).

Model Results

The full structural equation model, with the resulting standardized coefficients, appears in Figure 2. It includes 1 latent variable (L2 aptitude, $e_1$), 6 observed variables, and 16 error terms. The model includes direct effects and correlations among the variables.

### TABLE 1
Measures and Descriptive Statistics

<table>
<thead>
<tr>
<th>Measure</th>
<th>Items</th>
<th>Pts.</th>
<th>$M$</th>
<th>$SD$</th>
<th>Rel.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Rote Memory</td>
<td>88</td>
<td>88</td>
<td>59</td>
<td>5.52</td>
<td>0.91</td>
</tr>
<tr>
<td>Phonetic Coding Ability</td>
<td>80</td>
<td>80</td>
<td>47</td>
<td>7.34</td>
<td>0.79</td>
</tr>
<tr>
<td>Grammatical Sensitivity</td>
<td>24</td>
<td>24</td>
<td>15</td>
<td>5.19</td>
<td>0.83</td>
</tr>
<tr>
<td>Phonological Working Memory</td>
<td>48</td>
<td>48</td>
<td>39</td>
<td>7.10</td>
<td>0.88</td>
</tr>
<tr>
<td>Strategy Use</td>
<td>80</td>
<td>400</td>
<td>240</td>
<td>15.20</td>
<td>0.72</td>
</tr>
<tr>
<td>Motivation</td>
<td>38</td>
<td>190</td>
<td>143</td>
<td>17.56</td>
<td>0.91</td>
</tr>
<tr>
<td>Reading</td>
<td>60</td>
<td>60</td>
<td>49</td>
<td>4.44</td>
<td>na</td>
</tr>
<tr>
<td>Listening</td>
<td>60</td>
<td>60</td>
<td>45</td>
<td>3.05</td>
<td>na</td>
</tr>
<tr>
<td>Speaking</td>
<td>1</td>
<td>20</td>
<td>18</td>
<td>2.16</td>
<td>na</td>
</tr>
</tbody>
</table>

*The speaking test is an oral proficiency interview (OPI) and is scored holistically. Within this sample, all learners obtained a 10 (2+ on the Interagency Language Roundtable [ILR] scale) or a 20 (a 3 on the ILR scale).
represents the measures in Table 1, represented with rectangles). To examine the predictors of L2 learning, the proposed model was fitted to the data with listening, reading, and speaking test scores as dependent variables. The model was accepted and fit the data very well, $\chi^2(12, N = 96) = 7.6, p = .815$, CFI = 1.00, RMSEA = .00, RMSEA confidence interval = .00, .07.

In Figure 2, the model’s direct effects are shown as single-headed arrows and correlations are shown as double-headed ones. Also shown are standardized coefficients. Within this model, only one direct effect was significant at the .05 level: the effect of strategy use on reading ($r = -.25, p = .03$). Low strategy use significantly predicted higher L2 reading ability. Strategy use did not significantly predict listening or speaking ability. Learners’ L2 aptitude and motivation failed to predict significantly greater success in listening, reading, or speaking ability.

The model estimated the total variance explained in strategy use, motivation, listening, reading, and speaking (the squared multiple correlations provided in the AMOS output). Aptitude explained 9% of strategy use’s variance and 7% of motivation’s variance. (In other words, the error variance of strategy use is approximately 91% of the variance of strategy use itself, and the error variance of motivation is approximately 93%). The predictors (strategy use, aptitude, and motivation) explained 2% of listening’s variance, 3% of speaking’s, and 6% of reading’s.

**DISCUSSION**

This study used structural equation modeling to investigate the accuracy of a current understanding of L2 aptitude within the context of Chinese language acquisition. The model, diagrammed in Figure 2, included four cognitive aspects of L2 aptitude (rote memory, phonetic coding ability, grammatical sensitivity, and working memory) and two noncognitive variables (motivation and strategy use). Within this model, L2 aptitude was predicted to affect L2 learning directly. It was also hypothesized that L2 aptitude would indirectly affect L2 learning, that its effects would be mediated by motivation and strategy use. The model was found to be plausible (the data displayed a very good fit to the model), which allows me to discuss the individual effects the factors have on one another within the model. I first discuss the significant effect of strategy use on reading. After explaining why I examine other effects even though they were not significant, I discuss the (nonsignificant) effects of motivation on L2 learning. I then discuss two additional nonsignificant but interesting results concerning the variables that contribute to aptitude: the small effect that working memory had on aptitude and the negative effect that grammatical sensitivity had on aptitude. To conclude, I discuss the model’s overall represented construct of L2 aptitude and its effect on learning.

**The Effects of Strategy Use on L2 Learning**

The model showed only one statistically significant effect: Strategy use inversely affected success in reading ($r = -.25, p = .03$). In other words, when a learner’s SILL average rose by 1 standard deviation, his or her reading score fell by .25 standard deviations. (The absolute value of the standardized regression coefficient $r$ represents how much the independent variable affects the dependent variable or latent trait: A positive sign represents a direct relationship, and a negative sign represents an inverse relationship.)

Although this inverse relationship might initially seem surprising, it is consistent with past research. Ehrman and Oxford (1995) reported that less successful learners randomly use various strategies, while more successful learners systematically use specific strategies for specific tasks. This replicates aspects of a *curvilinear pattern* commonly found by L2 strategy researchers using the SILL: Advanced learners use fewer strategies than intermediate learners, advanced learners use a subset of their former strategies, and advanced learners use their select strategies in new, creative ways that relate to the complexities of advanced language learning (Green & Oxford, 1995; Hong–Nam & Leavell, 2006; Leaver, 2005; Oxford, 2011).8 Research focused on reading skills has also shown the importance of using an effective subset of strategies (Kember & Gow, 1994). Reliance on a few select strategies demonstrates automaticity in learning, which is necessary in upper level classes (Alyousef, 2005). So in a study examining reading in an upper level class, it is actually not surprising that the students who used fewer strategies were more successful at learning to read.

The inverse relationship between strategy use and success in reading makes particular sense in the context of Chinese language acquisition. The data showed that the students who use fewer strategies tend to be those with higher aptitude (aptitude inversely affects strategy use at $r = -.30, p = .58$). Aptitude is largely comprised of rote memory ($r = .97, p = .59$). And rote memory is critical to learning to read Chinese (Everson &
Ke, 1997; Hayden, 2005; Shen, 2004; Xiao, 2002). So the students who used fewer strategies may have been more successful at reading in part because they were the ones with better memories. They may not have needed to venture into other strategies because their rote-memory-based strategies worked particularly well.

Interestingly, strategy use had a much smaller effect on listening \((r = - .11, p = .35)\) and almost no effect on speaking \((r = .02, p = .85)\). These results lend empirical support to existing theory about strategy use. Researchers have speculated that strategies matter less for listening and speaking than for reading (Chamot, 2005; Farrell & Mallard, 2006) because in listening and speaking the importance of social and interactive skills overwhelms the effect of strategy use (Nakatani & Goh, 2007); reading, of course, does not require social interaction. Most empirical L2 strategy research has investigated the effect of strategies only on the skill of reading (Plonsky, 2011). Thus, this study provides evidence that strategy use may affect the various skills differently.

Nonsignificant Effects

Most of the effects in the model were not statistically significant. Although nonsignificant results are usually ignored, there are three reasons they are worth considering in this study. First, some statisticians contend that nonsignificant effects can be important (Valentine & Cooper, 2003). They point out that statistical significance tells us little about the “practical significance or relative impact” of the effect size, and should not be used as a standalone measure of how much the intervention ‘matters’” (Valentine & Cooper, p. 1, emphasis in original). Second, significance testing has been criticized as an arbitrary means of interpreting continuous data (Oswald & Plonsky, 2010; Plonsky, 2011; Plonsky & Gass, 2011; Schmidt, 1996). The data in this model are continuous rather than dichotomous.

Finally, it makes sense to consider some of the nonsignificant effects in this study because the correlations among some of the independent variables (known as multicollinearity) may have artificially reduced the significance of the effects. In this model, for example, rote memory was correlated with grammatical sensitivity \((r = .45, p = .00)\) and phonological working memory \((r = .23, p = .04)\). These correlated independent variables may be explaining the same parts of the variation in aptitude (Thayer, 1991), making the significance of their independent effects on aptitude misleadingly small. The explanatory power—the true significance—of that part of the variation in aptitude is, in effect, divided up among the correlated variables. It is also interesting because grammatical sensitivity has been shown to predict instructed language learning and also the implicit memorization of rule-based language, but not incidental language learning (see Robinson, 2007, p. 263). Thus, perhaps the results indicate that these advanced Chinese language learners learned a considerable amount incidentally, learning that did not depend on grammatical sensitivity. Similarly, strategy use was correlated with motivation \((r = .40, p = .00)\), reducing the significance of their independent effects, as well. For these reasons, significance testing should not be overemphasized in the interpretation of the model presented in Figure 2.

The Effects of Motivation on L2 Learning

As with strategy use, the data on motivation are largely consistent with prior theory and empirical research. The data showed that aptitude negatively affected motivation \((r = -.27, p = .58)\), which squares with Dörnyei’s (2005b) view that the less aptitude a learner has, the more motivation the learner needs. Motivation was positively correlated with strategy use \((r = .40, p = .00)\), as found by Vandergrift (2005). The data showed that motivation positively affects reading \((r = .16, p = .17)\), which is consistent with the theory that motivation is important to learning an L2 (Dörnyei, 2001a, 2005b; Ellis, 1994). (An effect size of .16 could be viewed as large within the context of advanced language learning.) While motivation ebbs and flows (Dörnyei, 2005b, 2009b), these data suggest that motivation has some enduring power even at the advanced level. One final point: Motivation had less of an effect on listening and speaking than on reading, just as strategy use did. As discussed above, the reason for the different effect of motivation on the different skills may be that gains in listening and speaking are more strongly influenced by socially construed interlocutor effects, diminishing the relative impact of motivation.

The Components of L2 Aptitude

One of the most important features of the model in Figure 2 is how it represents the construct of L2 aptitude. L2 aptitude is composed of four independent variables: rote memory, phonetic coding, grammatical sensitivity, and working memory. The regression coefficient (“r”) of each
reveals how much the score for aptitude is expected to increase when the score for the independent variable increases by 1 standard deviation, holding all other independent variables constant (Field, 2009; Thayer, 1991). In this model, aptitude for advanced-level Chinese consists of high rote memory, phonetic coding, and working memory abilities, and low grammatical sensitivity. Rote memory and grammatical sensitivity contribute the most to differences in aptitude. Increases in rote memory scores lead to major increases in aptitude scores \((r = .97, p = .59)\), and increases in grammatical sensitivity lead to major drops in aptitude scores \((r = -.86, p = .58)\). Increases in phonetic coding ability scores lead to much smaller increases in aptitude scores \((r = .14, p = .86)\), and increases in phonological working memory make almost no difference to aptitude scores \((r = .03, p = .92)\).

**Working Memory and L2 Aptitude.** It may seem difficult to reconcile the results regarding working memory and L2 learning when other studies have found working memory to be a significant predictor of L2 learning success (e.g., Harrington & Sawyer, 1992; Mackey et al., 2010; Mackey et al., 2002; Palladino & Cornoldi, 2004; Payne & Ross, 2005; Segalowitz & Lightbown, 1999; Tokowicz, Michael, & Kroll, 2004). However, as explained by Thayer (1991), “a variable might be the most important single predictor of a dependent variable when used alone but an unimportant predictor when used in combination with other predictors due to the amount of shared predicted variance” (p. 3). In this dataset, working memory correlates with rote memory \((r = .23, p = .04)\), which might account for working memory’s diminished role in defining the construct of L2 aptitude for Chinese. In any case, these results substantiate those from Hummel (2009), who found that working memory is not good at differentiating learning at upper levels of instruction. Unique in this study’s dataset is a potential reason why: There are other L2 aptitude factors that account for some variance in advanced L2 Chinese learning—the primary one being rote memory.

**Grammatical Sensitivity and L2 Aptitude.** A principal puzzle to solve is why, in this model, grammatical sensitivity inversely contributes to the construct of aptitude. When grammatical sensitivity (the ability to recognize the function of an English word in a sentence) goes down by 1 standard deviation, aptitude goes up by .86 standard deviations. It could be that students who understand English grammar well have difficulty adapting to the very different Chinese system of grammar. One could view those with high scores on grammatical sensitivity as those for whom the L1 system is firmly entrenched. That is, their perceptions of grammar may be tuned by the L1 to the extent that “their learned attention blocks them from perceiving differences in the L2” (Beckner et al., 2009, p. 10), which may, in this case, inhibit them, to some extent, from advanced Chinese language acquisition, or at least delay it. This finding is congruent with other research that has found that learners entrenched in their L1 patterns may have negative L1–L2 crosslinguistic influence, which might manifest itself during L2 acquisition by an overgeneralization of rules, avoidance of certain forms or structures, overproduction, and hypercorrection (Beckner et al., 2009; MacWhinney, 1997; Odlin, 1989). Aptitude test score interpretations as currently applied have not done well in predicting Chinese learning (Carroll, 1962, 1990, 1993). This study could be indicating why: According to this model, not all of the aptitude components are positively oriented for success in Chinese development.

Such information may have implications for pedagogy. Chinese instruction for native English speakers could focus more on non-L1-like, Chinese conceptual patterns and ways of thinking that are different from English, in particular, explicit instruction on L2 speech and writing patterns in Chinese that have been shown to be extremely difficult for nonnative speakers of English to acquire because they do not conform to any rules or sequential patterns present in English (Li, 2009, 2010). It has been proposed that profiling aptitudes could help match learners to instructional options and pedagogical tasks that would improve comprehension and production (Robinson, 2002a, 2002c). This study does so on a group level. The profile of aptitude for advanced-level Chinese reveals the great importance of rote memory and the necessity for openness to novel forms and ways of thinking about language and grammar.

**The Overall Effects of L2 Aptitude on L2 Chinese Learning.**

To finalize, as can be seen in Figure 2, no single component explains the observed variance in advanced L2 Chinese development. Within this model, aptitude, strategy use, and motivation have relatively the same impact on learning, as has been suggested by previous research (Dörnyei, 2001b; Schmidt & Watanabe, 2001; Tseng & Schmitt,
2008; Vandergrift, 2005). While strategy use and motivation mediate the effects of aptitude and have an influence on reading ability, they affect listening and speaking less. Aptitude, on the other hand, appears to directly affect speaking the most and does not directly affect reading or listening much at all. These seemingly discordant findings make intuitive sense when considering that language learning is a complex system affected by the interaction between the language learner and the language learning environment (Dörnyei, 2009a; Larsen–Freeman & Cameron, 2008). At the advanced stages of L2 learning, that interaction has had more time to exert its influence and creates unaccounted-for noise in the model. Within a dynamic systems view of L2 learning, high aptitude, high motivation, and good strategy use may be significantly advantageous conditions for attaining advanced proficiency, but when instruction is task-based and grounded in social interaction, minute distinctions in advanced proficiency may depend more on unmeasurable and unsystematic factors external to the model.

The predictor variables (aptitude, strategy use, and motivation) only explain 2%, 3%, and 6% of the variance in listening, speaking, and reading, respectively. These results may be disappointing for those looking for ways to predict which adults will successfully learn a foreign language to an advanced level of proficiency, but they are not surprising, Carroll (1962) found that after native-English-speaking learners of Chinese progressed beyond the beginning level, significant associations between aptitude (as measured by the MLAT) and learning became nonsignificant. Changes in Chinese L2 pedagogy since 1962 have made the question of what predicts advanced-level, Chinese-language-learning success worth asking again. Because working memory has been shown to be related to L2 learning success, it was sensible to include working memory as an additional component of L2 aptitude in such an investigation. Likewise, I included measures of strategy use and motivation because recent theorizing suggests the effects of aptitude are mediated by them. Based on the results of this study, it does not appear that aptitude, updated as a construct that includes working memory and two mediator variables (strategy use and motivation), is any better at explaining differences in advanced L2 Chinese attainment.

Why does L2 aptitude predict advanced-level L2 Chinese performance so poorly? What is it that distinguishes performance at the advanced level? I speculate that Larsen–Freeman and Cameron (2008) were right—the language learning environment is responsible for much of the learning that takes place. Different aptitude abilities are best suited to different stages of learning (Robinson, 2005a, 2007), and aptitude is not so important at the later stages of learning. Perhaps it is not that aptitude abilities are dynamic (as asserted by Skehan, 2002); rather, it might be that aptitude interacts with a changing learning environment (Neufeld, 1979; Snow, 1987; Sternberg, 2002; Sternberg & Grigorenko, 2002), thus bringing about flux (and a gradual reduction) in the importance of aptitude for learning. In sum, at the advanced level, it is not the cognitive or affective variables—the factors that lie within the learner (Ellis, 2004)—that matter; rather, it is how the individual reacts to the learning context. The learner’s actions in the social environment, the amount of time spent outside of class learning, and other personal reactions and choices of what to focus on ultimately affect learning.

LIMITATIONS AND FUTURE DIRECTIONS

An important limitation to this study is the time at which the motivation and strategy surveys were administered. Researchers have described how motivation changes over time for any given learner and have described how a flux in motivation may be related to temporal components as small as a task in the language learning classroom or as large as the flow of a foreign language course (Dörnyei, 2003, 2005b; Dörnyei & Ottó, 1998). Likewise, strategies that are useful at the beginning of language learning may not be those that are useful at the upper levels of language learning (Griffiths, 2003; Hong–Nam & Leavell, 2006); thus, strategy use, too, must naturally flux in response to the learning conditions. The motivation and strategy surveys were administered during the learners’ 42nd week of intense instruction, when most learners were most likely at an intermediate level of L2 Chinese. It could be that at this stage of such a lengthy process of highly controlled, classroom-based learning, motivation is fairly well established and generalized to a certain extent (Gardner et al., 1997; Tremblay & Gardner, 1995). Strategy use may reflect the learners’ individual proficiency levels at the time. The survey questions may have been replied to differently if administered to the same learners at a different point in their learning trajectory; thus, the results concerning motivation and strategies must be interpreted with the context in mind.
Another limitation is how I operationalized motivation and strategy use in this study. I created composite scores for these latent traits to keep the cases-per-parameter ratio low. According to Kline (2011), the number of participants to the number of free parameters should be 20 to 1 (yet 10 to 1 is perhaps more realistic for applied linguistics studies). By collapsing scores from the surveys, I asserted that the traits are one-dimensional. In reality, the motivation survey and the SILL do not each measure a homogeneous construct. Each trait comprises several subareas: motivation (as measured by the survey in this study; see Kormos & Dornyei, 2004) has seven; strategy use (as operationalized by the SILL, see Oxford, 1990a) has five. In future research, not to violate the assumption of composite score homogeneity, researchers could focus on certain areas of each latent trait. Or, to understand more effectively the relationship between, for example, strategy use and the different advanced skill areas (in this case, advanced listening, speaking, and reading), it would be useful to explore which questions best predict advanced-level success in the various skills. Currently, the items on the motivation survey and the SILL are not balanced by language learning skill areas (i.e., the SILL has very few items that address strategies for speaking) and are not normally segregated into skill areas. Nor do probabilistic statistics exist that explain which motivational aspects/strategies are most likely used by learners at certain levels of proficiency. (Such research would be interesting and useful.) Furthermore, the items are not streamlined as to which are most relevant for learning Chinese, a language proven difficult for English-language speakers because of its character-based writing system and use of tones. Thus, there is much groundwork to be done in terms of motivation and SILL instrument adaptation and refinement. Such research needs to be conducted to make data from the surveys more applicable to applied linguistics research using SEM.

In this study, the factor loading of working memory to aptitude is very low, which is inconsistent with the hypothesized model. This needs to be revisited. I assessed working memory through a single listening span test. It would be useful to administer different kinds of working memory tests because working memory tests, like the various subtests of L2 aptitude, may tap into different aspects of working memory. A more robust analysis of participants’ working memory skills could include reading span tests, visual–spatial tests, and perhaps even a nonsense word or digit span test, as recommended by Conway et al. (2005) and Waters and Caplan (1996). A more variegated measure of working memory may tell researchers more about this construct in relation to aptitude and the other variables in the model, but adding such parameters requires more participants.

Although there is substantial support for the final hypothesized model in this study in terms of goodness-of-fit indices, cross-validation SEM studies are needed to investigate whether the model in this study holds with other samples and whether those model estimates are stable across samples. Ideally, such studies would include a larger sample size: In this study, with 15 degrees of freedom and 96 participants, the power to reject the model is less than .2 (see Table 2 in MacCallum, Browne, & Sugawara, 1996). To achieve the statistically desired power of .8 (an 80% probability that the test would not make a Type II error, that is, fail to reject a null hypothesis when it is actually not true), 500 learners would be needed (MacCallum et al., 1996). Obtaining such participant numbers is notably difficult, especially when investigating advanced-level learners of a less-commonly-taught language such as Chinese. (For example, to obtain data from 500 participants for this study, I would have to collect data from Chinese language learners at the DLI for 5 years in a row; my 1 year of data collection yielded viable data from 96 participants at the cost of US$11,000 in National Science Foundation funding.) For this reason, testing the model or similar models out on diverse datasets may be more feasible. This needs to be done because we need to understand better how much variability is due to the learning context, the level of acquisition, and the specific L2 being learned. Future studies are also needed to explore the validity of using other kinds of working memory tests and other (perhaps more finely tuned) measures of motivation, strategy use, and L2 learning. Only then will we come closer to understanding the construct of L2 aptitude and how it relates to the complex system of L2 learning.

CONCLUSION

Several varied factors must successfully converge for an adult learner to obtain advanced proficiency in a foreign language. The learner needs excellent instruction, frequent opportunities for different kinds of output, and a heavy dose of motivation. Access to cultural insights that explain the pragmatics of the language has to come the learner’s way; effective language learning strategies need to be found; and, as any learner knows, real, tangible rewards for learning efforts must
materialize. It is natural that linguists search for a concrete reason why some succeed and others do not. But many individuals achieve advanced-level proficiency in a foreign language. And for any one individual working toward advanced proficiency, if his or her approach to learning changes regardless of his or her underlying cognitive strengths, advanced-level proficiency may come a few months earlier or later. Trying to pinpoint who will succeed fastest within a range of 3 to 6 months may be futile. Indeed, as evidenced by the data in this study, aptitude operationalized as a construct so extremely sensitive to such minute discrepancies in L2 learning may be unrealistic.

The results of this study provide useful information that helps us conceptualize the factors involved in the acquisition of advanced-level Chinese by adult native speakers of English. The results provide evidence that L2 aptitude, defined as rote memory, phonetic coding ability, grammatical sensitivity, and phonological working memory, is only a moderately useful construct in this context. Moreover, the effects of L2 aptitude on advanced-level learning are mediated by the affective variables of motivation and strategy use. In the 42nd week of a 63-week, intensive Chinese program, L2 aptitude predicts end-of-course performance no better than strategy use and motivation do. Variance in reading, listening, and speaking is only minimally explained by the predictors in this study’s model. Future research might focus on the observed, interactive, reciprocal aspects of motivation and strategy use; the differential effects aptitude, motivation, and strategy use have on the different skills of L2 development; or the relationships between rote memory and the other L2 aptitude factors observed in these data.

ACKNOWLEDGMENTS

I would like to thank the contributors to this research. The National Science Foundation funded this project (Award #0418175). Jeff Connor–Linton, Alison Mackey, and Charles Stansfield formed my dissertation committee at Georgetown University and guided this research. Gordon Jackson and John Lett at the Defense Language Institute provided logistical support and helped me solicit the Chinese-language-learner volunteers. Laura Klem (University of Michigan) and Alexander von Eye (Michigan State) provided technical feedback on the modeling. Comments from Helen Carpenter, Akiko Fujii Kurata, Heather Weger, and four anonymous MLJ reviewers assisted me in revising the paper, which I believe made it stronger. Any mistakes, however, are solely mine.

NOTES

1 Carroll envisioned a parallel “Form B” of the MLAT; however, Form B was never created. Nonetheless, the MLAT is still sold as “Form A.”
2 An additional factor, which Carroll identified as inductive language learning ability (the ability to decode linguistic material and conceptualize how other linguistic material would be encoded in the same language), was identified through tests that were acknowledged by Carroll to be administratively difficult; therefore, this construct, which Carroll (1962) stated was important for language learning, is not represented on the MLAT.
3 The DLAB is currently being revised. Go to CASL’s Web page (http://www.casl.umd.edu/dlab2) for more information. For current information on the PLAB and how it differs from the MLAT, go to the Language Learning and Testing Foundation’s Web pages on the PLAB (http://lltf.net/aptitude-tests/language-aptitude-tests/pimsleur-language-aptitude-battery) and on L2 aptitude testing in general (http://lltf.net/aptitude-tests/what-is-language-aptitude).
4 Traditional theories of working memory suggest that short-term memory works in combination with other factors within working memory (Baddeley, 2007; Daneman & Carpenter, 1980) and/or that information stored in long-term memory can be retrieved during working memory executive processes (Cowan, 2005). For more information, see Dehn (2008).
5 A few researchers posit that L2 aptitude is malleable. See information on and debates concerning Feuerstein’s Instrumental Enrichment, one purpose of which is to improve aptitude (Bailey & Pransky, 2010; Savell, Twohig, & Rachford, 1986).
6 Sasaki collected data from 160 Japanese students learning English in college in Japan. She used SEM to investigate the relationships among English proficiency, general intelligence, and L2 aptitude. But as in Hummel’s (2009) study, the quality and methods of the college students’ prior English language instruction were not controlled.
7 The measure was revised by increasing the number of sentences from 36 to 48 (Winke, Stafford, & Adams, 2003). The general procedures for this working memory span test are also reported in Mackey et al. (2010). The measure can be downloaded from the Instruments for Research into Second Languages (IRIS) database at http://www.iris-database.org.
8 I am extremely grateful to an anonymous MLJ reviewer who recommended this part of the discussion.
9 An anonymous MLJ reviewer noted that one way to measure the strategies of advanced language learners would be to interview them individually or in focus groups. The reviewer noted that advanced learners’ needs may be more idiosyncratic than what are measured by standardized surveys such as the SILL. Such information may help a researcher devise a skill- and language-specific Advanced Learners’ SILL.
REFERENCES


Masgoret, A.-M., & Gardner, R. C. (2003). Attitudes, motivation, and second language learning: A meta-


APPENDIX A

Motivation Questionnaire

This questionnaire is designed to gather information about how you, as a student, feel about learning Chinese. Please read each statement. Mark the response that tells how much you agree or disagree with each statement as follows: 1 = Strongly disagree, 2 = Disagree, 3 = Somewhat agree, 4 = Agree, 5 = Strongly agree

Part A.

1. Sometimes I feel that language learning is a burden for me.
2. I would like to get to know as many native speakers of the language I am learning as possible.
3. I am sure that I will be able to learn the language I am studying.
4. I think I am good at learning languages.
5. When I have to speak in my language class, I often lose confidence.
6. I like to work hard.
7. Unfortunately, I am not too good at learning the language I am studying.
8. I would rather spend time on subjects other than the language I am learning.
9. I am pleased with my current level of language ability in my language class.
10. I would like to spend a lot of energy learning this language in the future.
11. I am not too interested in my language class.
12. Learning this language often causes me a feeling of success.
13. In my parents’ view, the language class I am taking is not a very important course.
14. I would be pleased to be able to master an intermediate level of this language.
15. I really like the language I am learning.
16. I generally feel uneasy when I have to speak the language I am learning.
17. I generally feel uneasy when I have to read the language I am learning.
18. I generally feel uneasy when I have to write in the language I am learning.
19. We learn things in the language class that will be useful in the future.
20. Learning this language is one of the most important activities for me.
21. I rarely do more work for my language course than what is absolutely necessary.
22. I would like to get to an advanced level in this language.
23. I don’t mind it if I have to speak in this language with somebody.
24. I am satisfied with the work I do in my language class.
25. I easily give up the hard-to-reach goals.
26. I like my language class.
27. I would like to get to know many people who come from countries where this language is spoken.

Part B.
Learning this language is important to me…
1. … because I would like to get to know the culture and art of its speakers.
2. … because I may need it later for work or further education.

APPENDIX B

Pearson Product Moment Correlations between All Observed (Indicator) Variables

<table>
<thead>
<tr>
<th>Measure</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Rote Memory</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2. Phonetic Coding Ability</td>
<td>.12</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3. Grammatical Sensitivity</td>
<td>.45 * .01</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4. Phonological Working Memory</td>
<td>.23 ** .12 .09</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>5. Strategy Use</td>
<td>-.20</td>
<td>-.12</td>
<td>.10</td>
<td>-.13</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>6. Motivation</td>
<td>-.20</td>
<td>-.07</td>
<td>.09</td>
<td>-.05</td>
<td>.40 **</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>7. Listening</td>
<td>.20</td>
<td>.09</td>
<td>.15</td>
<td>.01</td>
<td>-.11</td>
<td>.02</td>
<td></td>
<td></td>
</tr>
<tr>
<td>8. Reading</td>
<td>.12</td>
<td>.04</td>
<td>.13</td>
<td>-.04</td>
<td>-.19</td>
<td>.07</td>
<td>.59 **</td>
<td></td>
</tr>
<tr>
<td>9. Speaking</td>
<td>.08</td>
<td>-.04</td>
<td>-.12</td>
<td>.03</td>
<td>-.06</td>
<td>-.10</td>
<td>.21</td>
<td>.32 **</td>
</tr>
</tbody>
</table>

Note. *p < .05, **p < .01.