RESEARCH

Language Aptitude and Crosslinguistic Influence in Initial L2 Learning

Lars Bokander

Language-learning aptitude and crosslinguistic similarity between learners' first language (L1) and the target second language (L2) are both known to facilitate successful L2 learning. However, these phenomena have rarely been investigated together in the same study. To address this research gap in second language acquisition, the present study was carried out with 92 international students of Swedish as a L2, with diverse L1 backgrounds. The participants first completed a language aptitude test upon entering a six-week introductory L2 course at the beginning level. Their L1 background was categorized in relation to the target language as either similar (Germanic L1) or distant (non-Germanic L1). At the end of the course, the participants completed a test of L2 achievement. Regression analyses of achievement scores, with language aptitude and L1 background as independent variables, revealed that crosslinguistic similarity explained at least as much variance in L2 achievement as did language aptitude. When comparing the effects of aptitude in the two L1 subsamples, language aptitude was found to be more important for the learners with a typologically similar L1, than for the learners with a more distant L1. In addition, the results provide support to theoretical proposals made in the individual differences literature that indicate that auditory processing ability may be of particular importance in the earliest stages of L2 acquisition.

1. Introduction

Second-language (L2) classrooms may be heterogeneous in many ways, for example with respect to the learners' age, aptitude, motivation, educational and linguistic background (Ellis, 2008). The variables of most interest in a given setting are the ones that are most clearly related to learning outcomes in that particular context. Hence, if all learners in a group are about the same age, or equally motivated to learn, individual differences in those variables tend to be small and inconsequential for learning outcomes. On the other hand, where differences in a variable are considerable, that variable may potentially explain important differences in language achievement. Foreign language aptitude and first-language (L1) typological proximity to the L2 to be learned are examples of variables whose impact on language acquisition has been shown to be considerable, but in distinctly separate research traditions (Li, 2019; Odlin, 1989). However, research on language aptitude has mostly relied on participant samples sharing the same L1, not permitting analysis of interactions between language background and aptitude. Similarly, research on crosslinguistic influence (CLI) on L2 learning has rarely taken language aptitude into consideration, because the predominant focus has been on differences in linguistic structures. Few, if any, studies have investigated interactions between, or the relative importance of, aptitude and CLI. It is largely unknown to what extent high language aptitude can compensate or offset any disadvantages brought about by a typologically distant L1.

Heterogenous, multilingual language-learning groups are commonly found in L2 classrooms around the world, for instance in language programmes for international students, adult immigrants or schools in linguistically diverse areas (García & Sylvan, 2011; Rosiers et al., 2016). As many language teachers can testify, students in L2 classrooms may progress at a highly different pace, sometimes to the extent that it is difficult to provide coherent instruction to the same group of students. It may thus be desirable to have some advance knowledge of which students have greater probability of a fast progression and which ones will need more time, support, and perhaps a different syllabus. Addressing the heterogeneous context faced by many L2 teachers in multilingual classrooms, this article aims at exploring contributions of both language learning aptitude and L1 background to successful L2 achievement, examining data from a mixed L1 group of adult beginning learners of Swedish as a L2.

Linnaeus University, SE
lars.bokander@lnu.se

1.1. Language learning aptitude

Among individual differences that influence (adult) L2 learning, language aptitude has played a prominent role in research ever since the first major language aptitude test battery was developed by Carroll and Sapon (1959). Along with motivation, language aptitude is generally considered to be the most determining individual-difference variable in L2 acquisition (Dörnyei & Ryan, 2015), consistently accounting for about 10–30% of the variance in L2 learning outcomes (Li, 2016). Language aptitude has been investigated in relation to a wide range of issues, such as L2 starting age (Abrahamsson & Hyltenstam, 2008; DeKeyser et al., 2000), L1 attrition (Bylund et al., 2010, 2012), instructed L2 learning (Saito, 2017; Yalçın et al., 2016) and in relation to other cognitive variables including motivation and intelligence (Gardner, 1986; Sasaki, 1999).

Arguably, researchers hold a consensus view that language aptitude is a multidimensional phenomenon, and different theoretical models have been suggested to explain the aptitude construct and its components. The most influential one has been the empirically derived four factor model (e.g., Carroll, 1981) in which language aptitude was proposed to consist of phonemic discrimination, rote memory, grammatical sensitivity and inductive learning ability. In recent years, research has been directed towards investigating possible distinctions between aptitudes for explicit and implicit learning (Granena, 2013b, 2019; Linck et al., 2013) and towards a growing awareness of the role of working memory as an important component of language aptitude (Linck et al., 2014; Wen et al., 2017). Others have conceptualized aptitude as the ability to handle novelty and ambiguity in language learning (Grigorenko et al., 2000), as essentially dependent upon first language (L1) ability (Sparks et al., 2011) or as aptitude complexes in interaction with the specific learning environment (Robinson, 2001).

Skehan (e.g., 2002, 2019) contributed important theoretical advancement by integrating aptitude research with processing stages in the field of second language acquisition (SLA), suggesting that different aptitude components are differentially important in the sequential stages of L2 development. Specifically, and with relevance for the present study, Skehan (2002) argued that the initial success of L2 beginners is particularly dependent upon phonological awareness (auditory processing), whereas language analytic ability (analogous to grammatical sensitivity and inductive ability in Carroll's, 1981, four factor model) becomes increasingly important at intermediate stages of L2 development. The reason for the special role of auditory processing at initial stages is its essential function in noticing and identifying patterns in the early L2 input. Besides phonological discrimination and encoding ability, auditory processing also includes the involvement of phonological short-term memory and executive working memory (Kormos, 2013). It is, however, less well known how these different components of auditory processing interact and how they are represented in language aptitude tests. The crucial role of auditory processing among beginning learners was supported empirically by Artieda and Muñoz (2016), who compared beginning and intermediate learners' L2 success in relation to LLAMA subtest scores (see Section 2.2 for details on the LLAMA test). They found a significant positive effect for sound sequence repetition ability with beginners but not with the intermediate learners, whereas in the latter group, the strongest correlation with L2 achievement was found with language analytic ability.

Several other classroom-based studies have investigated interactions between language aptitude (usually language analytic ability or working memory) and various treatment variables, including instructional method (Hwu et al., 2014; Li et al., 2019; Sanz et al., 2016), corrective feedback (Goo, 2012; Kourtali & Révész, 2020; Yilmaz & Granena, 2016) or complexity of grammatical structure (Yalçın & Spada, 2016; Yilmaz, 2013). Aptitude treatment interaction studies such as those mentioned have yielded mixed findings (see DeKeyser, 2019, for a general overview), but they mostly seem to suggest that language aptitude is helpful in situations that provide less explicit support for learners (e.g., under inductive instructional approaches that require learners to figure out grammatical rules without explicit presentation). Conversely, aptitude differences seem to become less important in deductive instructional approaches (rule presentation followed by practice) or when corrective feedback involves both recasts and metalinguistic explanation. Another finding (with particular relevance for the present study) is that difficult grammatical structures seem to require language analytical aptitude, whereas easier structures seem to be more memory dependent.

The cited studies on aptitude treatment interaction, together with research on aptitude effects related to other, non-manipulable and naturally occurring variables such as L2 proficiency level (Artieda & Muñoz, 2016), affect and motivation (Sparks et al., 2009) or age of acquisition (DeKeyser et al., 2010), constitute a wider class of research on individual-differences interactions in SLA (cf. DeKeyser, 2019). The present study extends that body of research by adding typological similarity to the list of aptitude interactions addressed so far.

1.2. Crosslinguistic influence

It is well testified that a learner's L1 influences the acquisition and use of a new language (e.g., Schepens et al., 2020). As a widely known example from outside academia, the Foreign Service Institute of the United States Department of State has for decades categorized languages in groups based on how long it is estimated that an English L1 speaker needs in order to acquire the L2 to a certain proficiency level (https://www.state.gov/foreign-language-training/). The role of the L1 in L2 learning is usually discussed in the SLA literature as transfer, or crosslinguistic influence (CLI). CLIs have been researched at all levels of language, with the strongest effects observed in phonology (foreign accent), but also in lexis, morphology, syntax, discourse, and pragmatics and in conceptual transfer (Jarvis & Pavlenko, 2008). A distinction has traditionally been made between positive (facilitation) and negative (interference) transfer, and, as pointed out by Jarvis and Pavlenko and Ringbom (1992; 2007), positive transfer (i.e., similarities between languages) appears to be

particularly important in initial L2 comprehension. One example is cognate facilitation (i.e., ease of comprehending historically related words with similar form in both L1 and L2), which has been extensively documented in psycholinguistic experimental research and in classroom-based, non-experimental studies (Helms-Park & Dronjic, 2016). Importantly, the presence of L2 cognates may allow learners to direct cognitive resources towards analysing other, more challenging parts of the L2 input, for example morphosyntactic structure (Ringbom, 2007). Research carried out in Finnish schools (where both Finnish and Swedish are spoken as L1s) with students learning English as a L2 (Jarvis & Odlin, 2000; Ringbom, 1992, 2007) has demonstrated significant advantages for the Swedish-L1 students compared to the Finnish-L1 students, due to the many lexical and grammatical similarities between English and Swedish (both Germanic languages, whereas Finnish belongs to the Finno-Ugric language group). The data analysed by Ringbom (2007) included student essays that provided global information from all language levels that may be observed in written form. Such an exploratory approach stands in contrast with the narrow tests of specific linguistic structures used in more experimentally oriented transfer studies (see the next paragraph).

Although CLIs occasionally have been studied in relation to individual variation in SLA (Odlin, 1989), very little research has been directed towards interactions between crosslinguistic similarity and individual differences in cognitive abilities. Effects of working memory on transfer in initial L2 learning were investigated by Trude and Tokowicz (2011) and Tolentino and Tokowicz (2014). The first of these studies examined pronunciation errors (i.e., negative transfer) by L1 speakers after a short training session in the L2 (a language previously unknown to the participants). The researchers found a negative correlation between working memory capacity and L1-like pronunciation errors, which they assumed to be due to stronger inhibition of the L1 in high working memory learners. Tolentino and Tokowicz (2014) compared learners' grammaticality judgements of L2 structures that differed in the degree of morphosyntactic overlap between the L1 and L2. Working memory did not interact significantly with type of structure in predicting learning outcomes, but an aptitude test of grammatical sensitivity (Carroll & Sapon, 1959) did, meaning that learners with strong grammatical sensitivity performed relatively better on structures that were dissimilar to the L1 or unique to the L2. The role of individual differences was, however, only a secondary aim in this study, and the combination of small sample size and complex research design suggests that power to detect interaction effects was low. The evidence for interactions between cognitive learning abilities and CLI is thus inconclusive, and more research is clearly needed in this area.

1.3. The present study

It is surprising that even though the topics of language aptitude and CLI are well-researched areas within SLA, they have only rarely been considered in the same study. Addressing this lacuna in SLA research, the present exploratory study, which was conducted longitudinally

at a Swedish university with international students, aimed at gauging the relative contributions of language aptitude and typological proximity to the target language (Swedish). Skill acquisition theory has suggested that in activities involving complex domain knowledge, prior knowledge is more important than differences in cognitive traits like working memory (Ackerman, 2007; see also DeKeyser, 2015; Sato & McDonough, 2019). This would indicate that learners who are able to draw on previous knowledge, in this case L1-L2 similarities, would have a greater advantage than learners who are high in cognitive skills but with no L1 similarities to guide their learning.

Furthermore, observing that language aptitude has previously been demonstrated to serve as a protection against negative effects of late starting age (e.g., DeKeyser, 2000), it may be hypothesized that aptitude plays a similar role in relation to CLI effects, such that it may offer some kind of protection against negative transfer, or that it will compensate for the absence of linguistic similarities between L1 and L2. If so, a larger effect of language aptitude would be expected among learners with a typologically distant L1.

Finally, because the learners in the study were absolute beginners, one may hypothesize, based on theory and previous empirical findings (Artieda & Muñoz, 2016; Skehan, 2002), that phonological aptitude would display a particularly strong association with L2 achievement. Because there is still a lack of empirical evidence within aptitude research validating the theory put forward by Skehan (2002), it was felt necessary to address this issue in the current study. Thus, the following research questions (RQs) were proposed.

> RQ1. What are the relative contributions of language aptitude and L1 background in a mixed-L1 group of L2 learners?
> RQ2. Will learners of typologically distant L1s benefit more from high aptitude, compared to those with a similar L1?
> RQ3. Do the present data confirm previous empirical findings demonstrating the theoretically expected importance of auditory processing in early L2 acquisition?

2. Method
2.1. Participants and learning context
The participants in the study were international university students (59 female and 33 male) studying Swedish at the beginner level. The majority were enrolled in computer science, business management, environmental studies or international relations. None of them was majoring in language or linguistics, but since the Swedish language course was optional and lessons were scheduled outside regular hours, it may be assumed that all were interested and motivated language learners. Their mean age was 23 years (SD = 4.5). L1s represented with at least five participants were: German (n = 28), English (n = 11), Japanese (n = 9), Mandarin (n = 9), Korean (n = 7), Dutch (n = 5), French (n = 5), Thai (n = 5). L1s with fewer participants were Arabic, Bengali, Czech, Farsi, Greek, Polish and Vietnamese. All knew English corresponding to at least

the upper-intermediate level. English was also an entry requirement at the Swedish university. All except four (native English speaking) participants reported knowing at least one more language besides their L1, at intermediate level or higher. No one reported any previous knowledge of Swedish or other Scandinavian languages.

In order to address the influence of L1 background in relation to aptitude, the participants were categorized as speakers of either typologically similar, or typologically distant L1s. Typologically similar was defined as having a L1 belonging to the Germanic language group (of which the target language, Swedish, is also a member). Typologically distant L1 was defined as any non-Germanic language (cf. Dryer & Haspelmath, 2013). This split produced two subsamples of roughly equal size – Germanic ($n = 44$) and non-Germanic ($n = 48$), arguably accounting for not just typological differences in, for example, lexis and morpho-syntactic structure, but also cultural/geographical differences that may impact CLIs (Ellis, 2008; Odlin, 1989).

The classroom instruction was characterized by a mix of communicative and form-focused approaches to teaching and learning, with two 3-hour sessions per week plus homework. The course syllabus emphasized spoken communication in everyday situations as a main objective, but also basic writing of simple sentences, and the textbook used in the course (targeted at level A1 + A2) highlighted communicative exercises and inductive learning of grammar as important features.

2.2. Materials and procedure

The participants completed a language aptitude test battery, and six weeks later, upon finishing the first course module, they completed a Swedish language proficiency test.

Due to the exploratory nature of the study, Swedish L2 proficiency was measured globally with a C-test (see Appendix), not targeting any particular linguistic structure. Language testing research has demonstrated the high validity of C-tests as measures of general language proficiency, tapping into textual, grammatical and lexical knowledge in both receptive and productive mode (Eckes & Grotjahn, 2006; Harsch & Hartig, 2016; Klein-Braley, 1997). For the purpose of this study, a C-test was constructed based on text excerpts from textbooks of Swedish for beginners. None of the textbooks sampled was used in the present language course. Four texts of about 100 words each were sampled and the second half of every other word was deleted, however leaving the first and last sentence intact. In total, there were 107 gaps in the texts, which had to be completed by the participant. The text excerpts were in the present tense and used short sentences with little or no subordination. Few of the target words required any inflection beyond their most frequent form. The C-test thus required very little grammatical analysis, but the participants needed to be able to understand the text and to recall enough vocabulary to complete the gaps.

The C-test was piloted with a small group of L2-learners and two native speakers. The students reported no inconsistencies and the native speakers both obtained full scores as expected (a valid C-test is normally solved without effort by educated native speakers). The test was then administered to the participating groups during their Swedish class and the session was supervised by the researcher or the teacher. There was no set time limit, but the testing time was practically limited to the remaining time of the lesson (about an hour, which was more than anyone needed). Scoring was done using the exact method, following the recommendation in Klein-Braley (1997). The internal consistency was high (KR-21 = 0.91), meaning that the test produced reliable scores (Crocker & Algina, 2008).

Language aptitude was measured with the LLAMA test battery (Meara, 2005), which has become a popular measure of language aptitude during the past decade. Several aspects of its validity have been investigated with speakers of a wide range of L1s (Bokander & Bylund, 2020; Granena, 2013a; Rogers et al., 2017) and, with importance for the present study, LLAMA scores do not seem to be influenced by the test takers' L1 background. Some of the LLAMA subtests have tended to produce low internal consistency estimates (typically resulting in wider confidence intervals for correlations with other variables, i.e., low power to detect aptitude related effects). However, in many studies (see Bokander & Bylund, 2020, for a comprehensive review), the LLAMA subtests have indeed demonstrated expected effects of language aptitude, thus implying that LLAMA may serve as a useful aptitude measure given that findings based on this instrument are interpreted with due carefulness.

The LLAMA test battery consists of four subtests, described in detail in Meara (2005). Each subtest begins with a timed exposure phase during which the test takers have to learn or remember some language material. This is followed by an untimed test phase during which the test takers respond to stimuli by selecting a response option on the computer screen. LLAMA B is a paired associates memory test where test takers have to learn the names of twenty pictures, which is followed by a pairing task. LLAMA D is a sound recognition test in which test takers have to decide whether a sound sequence is new to them, or previously encountered in the exposure phase. LLAMA E is a sound-symbol learning task, in which syllables are presented auditorily together with a new spelling system. In the test phase, test takers have to decide which is the correct spelling of two-syllable words that are played by the computer. Finally, in the LLAMA F exposure phase, test takers have to learn grammar and lexis by examining pictures and their descriptions in an unknown language. In the test phase, they choose which sentence correctly describes the actions in a presented picture. Maximum score was 100 in LLAMA B, E and F, and 75 in LLAMA D. It may be noted (with relevance for interpreting **Table 1** below) that the number of items in each LLAMA subtest are 20, except 30 in LLAMA D, but the reported total scores are formula scored percentages (cf. Frary, 1988, for an accessible account of formula scoring). Hence, a seemingly large difference in total score actually corresponds to a small difference in correctly answered items (e.g., one correct item is worth 10 percentage points in LLAMA E

Table 1: Mean (*SD*) scores in the Germanic and non-Germanic subsamples.

Variable	Germanic L1 (*n* = 44)	non-Germanic L1 (*n* = 48)
LLAMA B	54.0 (20.0)	56.0 (24.0)
LLAMA D	31.9 (14.4)	28.5 (18.0)
LLAMA E	65.0 (33.7)	73.8 (28.4)
LLAMA F	49.1 (30.2)	61.0 (25.3)
Swedish C-test	51.6 (15.8)	35.6 (17.8)

and F). The internal consistency (coefficient alpha) of the scores was 0.80 (LLAMA B), 0.55 (LLAMA D), 0.80 (LLAMA E) and 0.67 (LLAMA F). The lower alpha values for LLAMA D and F were expected, and in line with other studies that have used the LLAMA (for a comprehensive discussion of reliability issues in the LLAMA tests, see Bokander & Bylund, 2020).

2.3. Data analysis

In research designs combining categorical and continuous independent variables in non-experimental SLA settings (as the present study), where independent variables are frequently intercorrelated, multiple regression modelling has been advocated as a flexible method for handling these different kinds of intercorrelated data within the same analysis (Plonsky & Oswald, 2017). Because of the different metrics used in the independent variables, standardized regression coefficients were computed to ensure comparability between the independent variables. Also, the unique contribution of each variable to the regression model was computed as the incremental variance accounted for when entering that variable last in the equation (ΔR^2).

The analyses were carried out with the 'psych' package (Revelle, 2019) in R (R Core Team, 2018). Multiple linear regression models were computed with the *setCor* function in the 'psych' package.

3. Results
3.1. Summary statistics

In **Table 1** means and standard deviations are reported for each of the LLAMA subtests and the Swedish C-test. Pairwise *t*-tests revealed significant differences between the Germanic L1 and the non-Germanic L1 subsamples in LLAMA F (t(90) = −2.05, p = 0.044, d = −0.43, 95% CI [−0.84, −0.02]) and in the C-test (t(90) = 4.560, p < 0.001, d = 0.95, 95% CI [0.52, 1.38]). The effect size estimates (Cohen's *d*) indicate that the difference in LLAMA F scores may be considered small, whereas the difference in C-test scores were large, by the field specific benchmarks reported in Plonsky and Oswald (2014). In the other LLAMA subtests, any difference between the subsamples were statistically non-significant. It is thus evident that although the groups were similar in aptitude, the Germanic L1 subsample clearly outperformed the non-Germanic L1 subsample in Swedish proficiency.

Intercorrelations between the aptitude measures (**Table 2**) were small to medium (by field specific benchmarks in Plonsky & Oswald, 2014), suggesting that multicollinearity would not damage further analysis, although some covariance was involved between LLAMA B, E and F in both subsamples and also with LLAMA D and E in the non-Germanic subsample.

3.2. Regression models

Because some of the LLAMA subtest scores displayed moderate correlations between themselves, thus sharing variance, multiple regression was applied in order to find the unique contribution of each variable. Shapiro-Wilk tests confirmed that the residuals were normally distributed in all three regression models, at least W = .971, $ps \geq$ .284. The variance inflation factors (*VIF*) were well below 10 in all models, indicating that the standard errors of the regression coefficients were not unduly inflated by multicollinearity between independent variables (Pedhazur, 1997).

In the total sample (N = 92), the effect of L1 background was larger than any aptitude measure (**Table 3**). Altogether, the regression model accounted for 30% of the variance in C-test scores, F(5, 86) = 7.38, p < .001, R^2 = .30, 95% CI [0.15, 0.45]. The increment in explained variance (ΔR^2) upon entering Germanic L1 last in the regression equation was 0.15, meaning that L1 background explained about as much variance as the aptitude tests together. Importantly, however, the coefficients for all of the aptitude measures were of low significance, indicating that only the L1 effect may be considered reliable in this model. As evident upon examining **Tables 4** and **5**, the marginally significant effect of LLAMA D was completely driven by the Germanic subsample.

Table 2: Correlations (Pearson) between the LLAMA subtests in the Germanic and in the non-Germanic L1 subsamples.

Variable	LLAMA B	LLAMA D	LLAMA E	LLAMA F
LLAMA B	–	0.07	0.49*	0.46*
LLAMA D	0.17	–	0.26	0.11
LLAMA E	0.41*	0.42*	–	0.21
LLAMA F	0.27	0.12	0.44*	–

Note: Correlations in the Germanic L1 sample above the diagonal; correlations in the non-Germanic L1 sample below the diagonal. * = p < 0.05.

Table 3: Standardized regression model for the total sample (N = 92).

Variable	β	SE	t	p	ΔR^2	VIF
Germanic L1	.407	0.09	4.33	<.001	0.15	1.09
LLAMA B	.110	0.10	1.05	.295	0.01	1.33
LLAMA D	.191	0.10	1.98	.051	0.03	1.14
LLAMA E	.181	0.10	1.67	.099	0.02	1.44
LLAMA F	-.167	0.11	-1.66	.101	0.02	1.24

Table 4: Standardized regression model for the Germanic L1 subsample ($n = 44$).

Variable	β	SE	t	p	ΔR^2	VIF
LLAMA B	.139	0.16	0.847	.402	0.01	1.61
LLAMA D	.376	0.13	2.794	.008	0.13	1.08
LLAMA E	.325	0.15	2.121	.040	0.08	1.41
LLAMA F	-.193	0.15	-1.321	.194	0.03	1.28

Table 5: Standardized regression model for the non-Germanic L1 subsample ($n = 48$).

Variable	β	SE	t	p	ΔR^2	VIF
LLAMA B	.126	0.16	0.772	.444	0.02	1.21
LLAMA D	.132	0.16	0.809	.423	0.02	1.22
LLAMA E	.060	0.19	0.318	.752	0.01	1.64
LLAMA F	−.161	0.17	-0.973	.336	0.02	1.26

Because of the significant difference in mean C-tests scores between the similar and distant L1 groups, treating them as samples from the same population was not considered to be meaningful for the analysis of aptitude effects on L2 achievement, and thus separate regression models were computed (Pedhazur, 1997).

In the Germanic L1 subsample, the regression of C-test scores on the four aptitude tests (**Table 4**) explained 35% of the variance in the dependent variable, $F(4, 39) = 5.25$, $p = .002$, $R^2 = .35$, 95% CI [0.15, 0.55]. Significant effects were found for LLAMA D and E, that is, the two subtests in LLAMA tapping into auditory processing. Note that although the significant β-coefficients for LLAMA D and LLAMA E were of similar magnitude, the increment in explained variance (ΔR^2) was distinctly larger from LLAMA D due to its low correlation with the other subtests. No significant aptitude effects were found in the non-Germanic subsample (**Table 5**), $F(4, 43) = 0.65$, $p = .631$, $R^2 = .06$, 95% CI [–0.06, 0.18].

Summarizing the outcomes of the regression models, the answer to the first RQ is that L1 background was clearly more important than any language aptitude measure for explaining variation in L2 achievement (C-test scores). No significant effects of aptitude were found in the typologically distant, non-Germanic L1 subsample (**Table 5**), meaning that the hypothesis behind the second RQ was not supported. Possible reasons for this finding will be elaborated on in the next section of the paper. Finally, the significant effects of LLAMA D and LLAMA E in **Table 4** (the Germanic L1 subsample) constitute an affirmative answer to the third RQ, demonstrating the relative importance of auditory processing ability at the initial stage of L2 learning, at least for typologically similar languages.

4. Discussion

The present study aimed at gauging the relative contributions of L1 background and language aptitude to L2 achievement (measured as C-test scores) and at

exploring hypothesized differential effects of language aptitude on L2 learning, depending on the typological proximity between the learners' L1 and L2. A third aim was to investigate what aptitude dimension, as operationalized in LLAMA, seems to be most important at the very initial stages of L2 learning.

4.1. RQ 1: Relative contributions of language aptitude and L1 background

The first RQ addressed the relative contributions of language aptitude and of having a L1 that is similar to the target L2. It resembles questions raised elsewhere in the literature of individual differences, probing the roles of, for example, motivation or attitudes in relation to language aptitude as independent variables in SLA (Dörnyei & Ryan, 2015). For the first time, with this study, typological proximity has now been included in the family of relative-strength issues. Perhaps unsurprisingly, typological similarity to the target language demonstrated a large facilitating effect that, at least in initial stages of learning, seems to overshadow the benefits of high language aptitude (cf. Ringbom, 2007). As noted in the introduction of this paper, research drawing on skill acquisition theory has demonstrated how previously acquired domain knowledge has been found to be a better predictor of future skilled performance than cognitive abilities, such as working memory (Ackerman, 2007). Those findings are compatible with the results of the present study. Typologically related languages share many features, for instance, in phonology and vocabulary, and such features in the L1 (e.g., cognates) constitute previously acquired knowledge that may be transferred and utilized in L2 comprehension. Importantly, being able to take advantage of positive transfer in one aspect of the L2 means that attentional resources may be allocated to other aspects of the L2, which may increase the overall rate of learning (Ringbom, 2007). However, it may also be noted that once L1 background was controlled for by analysing L1-based subsamples separately, language aptitude accounted for more L2 variance in the Germanic L1-group (35%) than did L1 background and aptitude together in the more heterogeneous total sample (30%).

4.2. RQ 2: Different effects of language aptitude in typologically distant and similar L1 groups

The second RQ was formulated on the hypothesis that language aptitude would perform a similar function for speakers of distant L1s as it has been demonstrated to do for speakers with late age of onset of L2 acquisition (Abrahamsson & Hyltenstam, 2008). That is, the influence of aptitude on successful learning was supposed to be larger for speakers of dissimilar L1s than for speakers of similar L1s, just as the aptitude effect is stronger for post-adolescent L2 learners compared to young learners. However, the opposite relationship was observed. In the typologically similar, Germanic L1 group, a significant effect of (phonological) aptitude was detected, whereas no such effect was found in the typologically dissimilar, non-Germanic, L1 group. It thus appears as if late starting age and typological distance to the target L2 constitute two unrelated types of challenges for successful L2 learning.

What may then be the reason for the observed pattern, that language aptitude only seemed to have an impact in the typologically similar L1 group? A tentative explanation of the differential impact of aptitude in the two subsamples could draw on a two-way interaction between aptitude and transfer, such as suggested by Jarvis (2013) in the context of working memory and CLI. First, high aptitude may allow learners to capitalize on crosslinguistic similarities, whereas low aptitude learners can do that to a lesser degree. Capitalization on positive transfer is, however, possible only to the extent that similarities actually exist between L1 and L2. If the L2 is profoundly different from L1, there would be nothing to transfer even for learners with high aptitude. Second, typological similarity may permit immediate positive transfer and a head start in the L2 (as also discussed in relation to the first RQ), allowing for aptitude related processes to initiate, whereas aptitude without the help of initial positive transfer may need more time to take effect.

Finally, on the surface it may appear as if the lack of significant aptitude effects in the non-Germanic L1 group would be a statistical artefact due to a floor effect in the dependent variable (L2 proficiency), yielding less variance and hence less potential for detecting associations with the independent variables. However, the standard deviations of the C-test score distributions are about the same in both groups and there is thus substantial variance in learning outcome also in the non-Germanic group. That variability is obviously not captured by what the LLAMA measures. This invites the question of what kind of aptitude measure, if any, could have detected the variability in L2 achievement that obviously exists in the non-Germanic subsample. Perhaps some cognitive attributes not well represented in the LLAMA (e.g., aspects of working memory) or non-cognitive individual differences (e.g., personality or motivation) would provide better insights.

4.3. RQ3: The role of auditory processing in early L2 acquisition

The third RQ addressed what aptitude component would be most associated with L2 acquisition, possibly supporting theoretical proposals (e.g., Skehan, 2002, 2019; Wen et al., 2017) that auditory ability is relatively more important at the initial stages of L2 learning. The results in the present study align with previous findings by Artieda and Muñoz (2016) and with the theory of SLA processing stages (Skehan, 2002). In particular, LLAMA D and LLAMA E were significantly related to variance in Swedish achievement in the typologically similar L1 group. Both of these subtests involved listening skills – sound sequence recognition and sound-symbol pairings. Moreover, Granena (2013) demonstrated in a principal component analysis that of the four LLAMA subtests, LLAMA D displayed the highest loading on a component mainly associated with working memory. As noted in the introduction, a prominent role of working memory (including its phonological subsystem) in auditory processing has been hypothesized by several researchers (cf. Kormos, 2013; Skehan, 2019). It was noted in section

2.1. of this paper that the instructional approach of the Swedish course emphasized speaking skills. One may suspect that this learning context particularly favoured students with strong auditory processing ability.

Somewhat unexpected in this study was, however, the null findings for LLAMA F in both L1 groups. This subtest attempts to tap into language analytic ability, and although Skehan's (2002) theoretical framework predicts that the role of analytic ability increases after the earliest stages of L2 acquisition, several studies have reported positive correlations with analytic ability (or grammatical sensitivity) among beginners (e.g., Artieda & Muñoz, 2016; Tolentino & Tocowicz, 2014). One possible explanation for the lack of association between LLAMA F and learning outcomes in this study may lie in the way the L2 was assessed. For example, Tolentino and Tokowicz (2014) used a grammaticality judgement task that specifically tapped into the ability to analyse challenging morphosyntactic structures. The present study, on the other hand, used a global measure of L2 ability (C-test) in which the major challenge presumably was not morphosyntactic analysis, but rather, overall comprehension and retrieval of lexical items which only had to appear in their most frequent form to be scored as correct (Section 2.2).

5. Conclusion

Although valuable conclusions may be drawn from the present study, recognizing its shortcomings is due. First, the sample size was rather small for performing multiple regression analyses, albeit acceptable using more liberal recommendations (Pedhazur, 1997). This implies that measurement errors could be larger than desired, and as a consequence, power to detect true effects may be on the low side. Second, measurement errors may also have been increased by low internal consistency in the scores from LLAMA D and LLAMA F in particular. As noted in Section 2, this seems to be an inherent problem with the LLAMA, but with several studies reporting similar findings (e.g., the present study and Artieda & Muñoz, 2016), measurement imprecision may to some extent be counteracted. A way to handle low reliability in future studies could be to develop adapted test versions by adding more items, since this is a well-known method to increase reliability. In Suzuki and DeKeyser (2017), ten additional items were added to LLAMA F, resulting in a substantial increase in reliability over what other studies have reported with this test. Third, by using a convenience sample, important nuances in the effects of crosslinguistic similarity may be lost. Ideally, a replication study would involve a larger sample and recruiting different L1s, situated along a continuum of typological proximity to the L2. In that way, a more fine-grained analysis would be feasible. Finally, there is a possibility that the participants' knowledge of other L2s could play a role in how transfer and aptitude effects interact. All of the (non-native English speaking) participants in this study had at least upper intermediate knowledge of L2 English, but there could still be substantial differences in their English skills. If possible, future research should control for such variation.

Despite these shortcomings, some important insights have emerged. For the first time, the relative impact of language aptitude and L1-L2 similarity on L2 acquisition has been quantified. This, together with similar future findings, could have implications for the organization of L2 instruction in linguistically diverse contexts, recognizing that at least initially, high aptitude is not likely to compensate speakers of typologically highly dissimilar languages. However, it was also observed that once learners have been able to take advantage of positive transfer, there was an obvious effect of phonological aptitude on L2 learning (supporting theoretical expectations). Thus, in contexts where all learners in a group are (almost) homogeneous with respect to L1, taking (phonological) aptitude in consideration seems well motivated. An interesting question for future research would be to find out if there is a point in time during L2 development, after which language aptitude supersedes L1 background as the more influential individual-difference variable. Another important question concerns the generalizability to learners of different educational backgrounds. In the present study, all participants were university students. Future research replicating the study in other learning contexts (e.g., with immigrants of variable socio-economic situation) would be highly desirable. Finally, as also suggested by Jarvis (2013), future research may want to specify exactly what features of the L1 may transfer and investigate if language aptitude may influence different kinds of transfer in different ways. There are certainly many new phenomena awaiting to be discovered in the intriguing and under-researched area of language aptitude and transfer interactions.

Additional File

The additional file for this article can be found as follows:

- **Appendix.** C-test. DOI: https://doi.org/10.22599/jesla.69.s1

References

Abrahamsson, N., & Hyltenstam, K. (2008). The robustness of aptitude effects in near-native second language acquisition. *Studies in Second Language Acquisition*, *30*, 481–509. DOI: https://doi.org/10.1017/S027226310808073X

Ackerman, P. L. (2007). New developments in understanding skilled performance. *Current Directions in Psychological Science*, *16*, 235–239. DOI: https://doi.org/10.1111/j.1467-8721.2007.00511.x

Artieda, G., & Muñoz, C. (2016). The LLAMA tests and the underlying structure of language aptitude at two levels of foreign language proficiency. *Learning and Individual Differences*, *50*(Supplement C), 42–48. DOI: https://doi.org/10.1016/j.lindif.2016.06.023

Bokander, L., & Bylund, E. (2020). Probing the internal validity of the LLAMA language aptitude tests. *Language Learning*, *70*, 11–47. DOI: https://doi.org/10.1111/lang.12368

Bylund, E., Abrahamsson, N., & Hyltenstam, K. (2010). The role of language aptitude in first language attrition: the case of pre-pubescent attriters. *Applied Linguistics*, *31*, 443–464. DOI: https://doi.org/10.1093/applin/amp059

Bylund, E., Abrahamsson, N., & Hyltenstam, K. (2012). Does first language maintenance hamper nativelikeness in a second language? A study of ultimate attainment in early bilinguals. *Studies in Second Language Acquisition*, *34*, 215–241. DOI: https://doi.org/10.1017/S0272263112000034

Carroll, J. B. (1981). Twenty-five years of research on foreign language aptitude. In K. C. Diller (Ed.), *Individual differences and universals in language learning aptitude* (pp. 83–118). Newbury House.

Carroll, J. B., & Sapon, S. M. (1959). *Modern language aptitude test.* New York, NY: Psychological Corporation.

Crocker, L. M., & Algina, J. (2008). *Introduction to classical and modern test theory*. Mason, OH: Cengage Learning.

DeKeyser, R. M. (2000). The robustness of critical period effects in second language acquisition. *Studies in Second Language Acquisition*, *22*, 499–533. DOI: https://doi.org/10.1017/S0272263100004022

DeKeyser, R. M. (2015). Skill acquisition theory. In B. VanPatten & J. Williams (Eds.), *Theories in second language acquisition: An introduction* (2nd ed., pp. 94–112). London: Routledge.

DeKeyser, R. M. (2019). Introduction to the special issue of aptitude-treatment interaction in second language learning. *Journal of Second Language Studies*, *2*, 165–168. DOI: https://doi.org/10.1075/jsls.00007.int

DeKeyser, R., Alfi-Shabtay, I., & Ravid, D. (2010). Cross-linguistic evidence for the nature of age effects in second language acquisition. *Applied Psycholinguistics*, *31*, 413–438. DOI: https://doi.org/10.1017/S0142716410000056

Dörnyei, Z., & Ryan, S. (2015). *The Psychology of the Language Learner Revisited*. London: Routledge. DOI: https://doi.org/10.4324/9781315779553

Dryer, M. S., & Haspelmath, M. (Eds.). (2013). *WALS Online*. Max Planck Institute for Evolutionary Anthropology. https://wals.info/

Eckes, T., & Grotjahn, R. (2006). A closer look at the construct validity of C-tests. *Language Testing*, *23*, 290–325. DOI: https://doi.org/10.1191/0265532206lt330oa

Ellis, R. (2008). *The study of second language acquisition* (2nd ed.). Oxford: Oxford University Press.

Frary, R. B. (1988). Formula scoring of multiple-choice tests (correction for guessing). *Educational Measurement: Issues and Practice*, *7*(2), 33–38. DOI: https://doi.org/10.1111/j.1745-3992.1988.tb00434.x

García, O., & Sylvan, C. E. (2011). Pedagogies and practices in multilingual classrooms: singularities in pluralities. *The Modern Language Journal*, *95*, 385–400. DOI: https://doi.org/10.1111/j.1540-4781.2011.01208.x

Gardner, R. C. (1986). *Social psychology and second language learning: The role of attitudes and motivation.* Baltimore, MD: Hodder Arnold.

Granena, G. (2013a). Cognitive aptitudes for second language learning and the LLAMA language aptitude

test. In G. Granena & M. Long (Eds.), *Sensitive periods, language aptitude, and ultimate L2 attainment* (pp. 105–130). Amsterdam: John Benjamins. DOI: https://doi.org/10.1075/lllt.35.04gra

Granena, G. (2013b). Individual differences in sequence learning ability and second language acquisition in early childhood and adulthood. *Language Learning, 63,* 665–703. DOI: https://doi.org/10.1111/lang.12018

Granena, G. (2019). Cognitive aptitudes and l2 speaking proficiency: links between LLAMA and Hi-LAB. *Studies in Second Language Acquisition, 41,* 313–336. DOI: https://doi.org/10.1017/S0272263118000256

Goo, J. (2012). Corrective feedback and working memory capacity in interaction-driven l2 learning. *Studies in Second Language Acquisition, 34,* 445–474. DOI: https://doi.org/10.1017/S0272263112000149

Grigorenko, E. L., Sternberg, R. J., & **Ehrman, M. E.** (2000). A theory-based approach to the measurement of foreign language learning ability: The CANAL-F theory and test. *Modern Language Journal, 84,* 390–405. DOI: https://doi.org/10.1111/0026-7902.00076

Harsch, C., & **Hartig, J.** (2016). Comparing C-tests and Yes/No vocabulary size tests as predictors of receptive language skills, *Language Testing, 33,* 555–575. DOI: https://doi.org/10.1177/0265532215594642

Helms-Park, R., & **Dronjic, V.** (2016). Crosslinguistic lexical influence: cognate facilitation. In R. Alonso Alonso (Ed.), *Crosslinguistic influence in second language acquisition* (pp. 71–92). Bristol, UK: Multilingual Matters. DOI: https://doi.org/10.21832/9781783094837-007

Hwu, F., Pan, W., & **Sun, S.** (2014). Aptitude-treatment interaction effects on explicit rule learning: A latent growth curve analysis. *Language Teaching Research, 18,* 294–319. DOI: https://doi.org/10.1177/1362168813510381

Jarvis, S. (2013). Crosslinguistic influence and multilingualism. In C. A. Chapelle (Ed.), *The Encyclopedia of Applied Linguistics.* Malden, MA: Wiley-Blackwell. DOI: https://doi.org/10.1002/9781405198431.wbeal0291

Jarvis, S., & **Odlin, T.** (2000). Morphological type, spatial reference, and language transfer. *Studies in Second Language Acquisition, 22,* 535–556. DOI: https://doi.org/10.1017/S0272263100004034

Jarvis, S., & **Pavlenko, A.** (2008). *Crosslinguistic influence in language and cognition.* London: Routledge. DOI: https://doi.org/10.4324/9780203935927

Klein-Braley, C. (1997). C-tests in the context of reduced redundancy testing: An appraisal. *Language Testing, 14,* 47–84. DOI: https://doi.org/10.1177/026553229701400104

Kormos, J. (2013). New conceptualizations of language aptitude in second language attainment. In G. Granena & M. Long (Eds.), *Sensitive periods, language aptitude, and ultimate L2 attainment* (pp. 131–152). Amsterdam: John Benjamins. DOI: https://doi.org/10.1075/lllt.35.05kor

Kourtali, N.-E., & **Révész, A.** (2020). The roles of recasts, task complexity, and aptitude in child second language

development. *Language Learning, 70,* 179–218. DOI: https://doi.org/10.1111/lang.12374

Li, S. (2016). The construct validity of language aptitude. *Studies in Second Language Acquisition, 38,* 801–842. DOI: https://doi.org/10.1017/S027226311500042X

Li, S. (2019). Six decades of language aptitude research: A comprehensive and critical review. In Z. Wen, P. Skehan, A. Biedron, S. Li, & R. L. Sparks (Eds.), *Language aptitude. Advancing theory, testing, research and practice* (pp. 78–96). Routledge. DOI: https://doi.org/10.4324/9781315122021-5

Li, S., Ellis, R., & **Zhu, Y.** (2019). The associations between cognitive ability and L2 development under five different instructional conditions. *Applied Psycholinguistics, 40,* 693–722. DOI: https://doi.org/10.1017/S0142716418000796

Linck, J. A., Hughes, M. M., Campbell, S. G., Silbert, N. H., Tare, M., Jackson, S. R., Smith, B. K., Bunting, M. F., & **Doughty, C. J.** (2013). Hi-LAB: A new measure of aptitude for high-level language proficiency. *Language Learning, 63,* 530–566. DOI: https://doi.org/10.1111/lang.12011

Linck, J. A., Osthus, P., Koeth, J. T., & **Bunting, M. F.** (2014). Working memory and second language comprehension and production: A meta-analysis. *Psychonomic Bulletin & Review, 21,* 861–883. DOI: https://doi.org/10.3758/s13423-013-0565-2

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

Odlin, T. (1989). *Language transfer. Cross-linguistic influence in language learning.* NY: Cambridge University Press. DOI: https://doi.org/10.1017/CBO9781139524537

Pedhazur, E. J. (1997). *Multiple regression in behavioral research* (3rd ed.). Wadsworth, OH: Thomson Learning.

Plonsky, L., & **Oswald, F. L.** (2014). How big is "big"? interpreting effect sizes in L2 research. *Language Learning, 64,* 878–912. DOI: https://doi.org/10.1111/lang.12079

Plonsky, L., & **Oswald, F. L.** (2017). Multiple regression as a flexible alternative to ANOVA in L2 research. *Studies in Second Language Acquisition, 39,* 579–592. DOI: https://doi.org/10.1017/S0272263116000231

R Core Team. (2018). R: A language and environment for statistical computing. Retrieved from https://www.R-project.org

Revelle, W. (2019). *psych: Procedures for Psychological, Psychometric, and Personality Research.* Evanston, Illinois: Northwestern University. R package version 1.9.12, https://CRAN.R-project.org/package=psych.

Ringbom, H. (1992). On L1 transfer in L2 comprehension and L2 production. *Language Learning, 42,* 85–112. DOI: https://doi.org/10.1111/j.1467-1770.1992.tb00701.x

Ringbom, H. (2007). *Cross-linguistic similarity in foreign language learning.* Clevedon, UK: Multilingual Matters. DOI: https://doi.org/10.21832/9781853599361

Robinson, P. (2001). Individual differences, cognitive abilities, aptitude complexes and learning

conditions in second language acquisition. *Second Language Research, 17*, 368–392. DOI: https://doi.org/10.1177/026765830101700405

Rogers, V., Meara, P., Barnett-Legh, T., Curry, C., & Davie, E. (2017). Examining the LLAMA aptitude tests. *Journal of the European Second Language Association, 1*, 49–60. DOI: https://doi.org/10.22599/jesla.24

Rosiers, K., Willaert, E., Avermaet, P. V., & Slembrouck, S. (2016). Interaction for transfer: Flexible approaches to multilingualism and their pedagogical implications for classroom interaction in linguistically diverse mainstream classrooms. *Language and Education, 30*, 267–280. DOI: https://doi.org/10.1080/09500782.2015.1117097

Saito, K. (2017). Effects of sound, vocabulary, and grammar learning aptitude on adult second language speech attainment in foreign language classrooms. *Language Learning, 67*, 665–693. DOI: https://doi.org/10.1111/lang.12244

Sanz, C., Lin, H.-J., Lado, B., Stafford, C. A., & Bowden, H. W. (2016). One size fits all? learning conditions and working memory capacity in *Ab Initio* language development. *Applied Linguistics, 37*, 669–692. DOI: https://doi.org/10.1093/applin/amu058

Sasaki, M. (1999). *Second language proficiency, foreign language aptitude, and intelligence: quantitative and qualitative analyses* (2nd ed.). New York, NY: Peter Lang.

Sato, M., & McDonough, K. (2019). Practice is important but how about its quality? Contextualized practice in the classroom. *Studies in Second Language Acquisition, 41*, 999–1026. DOI: https://doi.org/10.1017/S0272263119000159

Schepens, J., van Hout, R., & Jaeger, T. F. (2020). Big data suggest strong constraints of linguistic similarity on adult language learning. *Cognition, 194*, 1–14. DOI: https://doi.org/10.1016/j.cognition.2019.104056

Skehan, P. (2002). Theorising and updating aptitude. In P. Robinson (Ed.), *Language learning & language teaching* (pp. 69–93). Amsterdam: John Benjamins. DOI: https://doi.org/10.1075/lllt.2.06ske

Skehan, P. (2019). Language aptitude implicates language and cognitive skills. In Z. Wen, P. Skehan, A. Biedron, S. Li & R. L. Sparks (Eds.), *Language aptitude: Advancing theory, testing, research and practice* (pp. 56–77). London: Routledge. DOI: https://doi.org/10.4324/9781315122021-4

Sparks, R. L., Humbach, N., Patton, J., & Ganschow, L. (2011). Subcomponents of second-language aptitude and second-language proficiency. *The Modern Language Journal, 95*, 253–273. DOI: https://doi.org/10.1111/j.1540-4781.2011.01176.x

Sparks, R. L., Patton, J., Ganschow, L., & Humbach, N. (2009). Long-term relationships among early first language skills, second language aptitude, second language affect, and later second language proficiency. *Applied Psycholinguistics, 30*, 725–755. DOI: https://doi.org/10.1017/S0142716409990099

Suzuki, Y., & DeKeyser, R. M. (2017). The interface of explicit and implicit knowledge in a second language: Insights from individual differences in cognitive aptitudes. *Language Learning, 67*, 747–790. DOI: https://doi.org/10.1111/lang.12241

Tolentino, L. C., & Tokowicz, N. (2014). Cross-language similarity modulates effectiveness of second language grammar instruction. *Language Learning, 64*, 279–309. DOI: https://doi.org/10.1111/lang.12048

Trude, A. M., & Tokowicz, N. (2011). Negative transfer from Spanish and English to Portuguese pronunciation: The roles of inhibition and working memory. *Language Learning, 61*, 259–280. DOI: https://doi.org/10.1111/j.1467-9922.2010.00611.x

Wen, Z. (Edward), Biedroń, A., & Skehan, P. (2017). Foreign language aptitude theory: Yesterday, today and tomorrow. *Language Teaching, 50*, 1–31. DOI: https://doi.org/10.1017/S0261444816000276

Yalçın, Ş., Çeçen, S., & Erçetin, G. (2016). The relationship between aptitude and working memory: An instructed SLA context. *Language Awareness, 25*, 144–158. DOI: https://doi.org/10.1080/09658416.2015.1122026

Yalçın, Ş., & Spada, N. (2016). Language aptitude and grammatical difficulty: An EFL Classroom-Based Study. *Studies in Second Language Acquisition, 38*, 239–263. DOI: https://doi.org/10.1017/S0272263115000509

Yilmaz, Y. (2013). Relative effects of explicit and implicit feedback: The role of working memory capacity and language analytic ability. *Applied Linguistics, 34*, 344–368. DOI: https://doi.org/10.1093/applin/ams044

Yilmaz, Y., & Granena, G. (2016). The role of cognitive aptitudes for explicit language learning in the relative effects of explicit and implicit feedback. *Bilingualism: Language and Cognition, 19*, 147–161. DOI: https://doi.org/10.1017/S136672891400090X