The Role of Language-Analytic Ability in Children's Instructed Second Language Learning

Karen Roehr-Brackin and Angela Tellier

University of Essex

kroehr@essex.ac.uk

atelli@essex.ac.uk

Department of Language and Linguistics

University of Essex

Wivenhoe Park

Colchester CO4 3SQ

United Kingdom

The Role of Language-Analytic Ability in Children's Instructed Second Language Learning

Abstract

Language-analytic ability, or the ability to treat language as an object of analysis and arrive at linguistic generalizations, is at the core of the constructs of language learning aptitude and metalinguistic awareness, which are implicated in our ability to learn explicitly. In the context of child second language (L2) learning, it has been argued that children learn primarily implicitly and that the most important component of aptitude may be memory ability. However, no empirical research to date has investigated the relationship and development of aptitude and metalinguistic awareness longitudinally as well as examined their predictive power for children's L2 achievement in the classroom. In a study with English-speaking learners aged 8-9 (N = 111), we found that although aptitude and metalinguistic awareness were (still) dynamic, they significantly predicted children's achievement in L2 French. Moreover, language-analytic ability proved to be the component with the strongest predictive power. This finding suggests that it may not be level of cognitive maturity alone that determines children's approach to L2 learning; experiencing explicit, form-focused instruction may foster the role of language-analytic ability even in children as young as 8-9 years.

Keywords: language-analytic ability, metalinguistic awareness, aptitude, form-focused instruction, child second language learning

The Role of Language-Analytic Ability in Children's Instructed Second Language Learning

Introduction

Teaching young children a foreign or second language (L2) has become the norm in many countries. Although research in naturalistic settings indicates that an early start to L2 learning will convey advantages in ultimate achievement compared with a later start (e.g. Johnson & Newport, 1989; DeKeyser, 2000), research in instructed contexts has failed to identify advantages for young learners. Indeed, after the same amount of classroom input, late starters have been shown to consistently outperform early starters on practically all L2 measures employed (e.g. García Mayo, 2003; Jaekel, Schurig, Florian, & Ritter, 2017; Larson-Hall, 2008; Muñoz, 2006, 2008, 2009; see also Qureshi, 2016 for a meta-analysis). This pattern of findings is often explained with reference to young children's reliance on primarily implicit learning mechanisms, which require both extensive and intensive long-term input to be successful, and adolescents' and adults' more developed capacity to draw on explicit learning processes, which are cognitively resource-intensive, but also potentially fast and efficient and thus especially useful in the limited-input setting of a typical L2 classroom. The success or otherwise of explicit learning in particular has been linked with individual learner differences, especially language learning aptitude and metalinguistic awareness (Roehr-Brackin, 2015, 2018; Tellier & Roehr-Brackin, 2017). As children mature and gradually develop the ability to learn explicitly, the question arises to what extent language learning aptitude and metalinguistic awareness play a role in their instructed L2 learning. The present paper addresses the conceptual question of how these two constructs relate to each other, the empirical question of if and how these abilities develop as children mature, and the practical question as to what extent their constituent components can predict children's achievement in the language classroom.

Theoretical and empirical background: Language learning aptitude and metalinguistic awareness

In the classic conceptualization of language learning aptitude (Carroll, 1962, 1981, 1990), the construct has four components: phonetic coding ability, or the ability to identify sounds in the foreign language; grammatical sensitivity, or the ability to recognize how words function grammatically in sentences; inductive language learning ability, or the ability to induce grammatical rules from language examples, i.e. identifying patterns of correspondence and form-meaning relationships; and associative memory, or the ability to recognize and remember words and phrases. More recently, researchers have conventionally subsumed grammatical sensitivity and inductive language learning ability under the label of language-analytic ability, defined as the ability to infer linguistic systematicities from the input and make generalizations (Skehan, 1998, 2002). The most commonly used measure associated with the four-component model of aptitude is the Modern Language Aptitude Test (MLAT: Carroll & Sapon, 1959; Carroll & Sapon, 2002b) for adult learners and the Modern Language Aptitude Test-Elementary (MLAT-E: Carroll & Sapon, 2002a) for child learners.

Language learning aptitude has proved a reliable predictor of achievement in instructed adult L2 learning (Dörnyei & Skehan, 2003; Li, 2016), but evidence pertaining to the role of aptitude in child L2 learning, defined here as foreign or second language learning up to the age of 12, is still in relatively short supply. Contrary to the findings of earlier work (e.g. DeKeyser, 2000), more recent studies examining L2 speakers with different starting ages have led to the conclusion that aptitude seemingly plays a role for both adults and children in naturalistic contexts (Abrahamsson & Hyltenstam, 2008; Granena, 2013, 2014), although the presence and/or magnitude of aptitude effects may depend in part on exactly how the construct is operationalized.

Studies with child L2 learners in classroom settings likewise indicate a role for aptitude (Kiss, 2009; Tellier & Roehr-Brackin, 2013). Working with 12-year-old Hungarian learners of L2 English (N = 419), Kiss and Nikolov (2005) administered an aptitude test modelled on the MLAT, and a proficiency test assessing English listening, reading and writing skills. Aptitude and L2 proficiency scores were significantly correlated at a medium level of strength, suggesting that the aptitude measure can serve as a predictor in 12-year-olds. Interestingly, the researchers argue that as no association between time spent learning the L2 and aptitude scores was found, aptitude appears to be fairly stable in 12-year-olds, with no evident improvement over time and/or through L2 exposure.

In a subsequent study, Kiss (2009) outlines the development of a test of language learning aptitude that relies on the MLAT-E and the MLAT as models. Ninety-two Hungarian children in two primary schools were tested to select 26 eight-year-olds most likely to cope with a newly established bilingual English-Hungarian teaching program. When results from the 8-year-old children were compared with those from 12-year-olds in a previous study, it became clear that the 12-year-olds did much better than the 8-year-olds on the vocabulary learning subtest. This finding is attributed to the older children's language learning experience and associated strategies that would have developed as a consequence, so the researcher acknowledges dynamicity in children's aptitude at least up to the age of 12.

After one year, the children in the bilingual program were tested for English oral proficiency by means of a 5-minute interview comprising a short warm-up, simple personal information questions, and a picture-based spot-the-difference task. Two markers scored the children's performance and two teachers ranked the children according to their overall in-class L2 performance. While ratings and rankings were correlated, only few significant correlations between aptitude scores and oral proficiency in English were found. Aptitude correlated with the oral marks awarded by one marker, while performance on one subtest correlated with the oral marks awarded by one marker and both teachers' rankings (Kiss, 2009).

It is not entirely clear how to interpret these results. First, only those children who demonstrated high levels of aptitude prior to the start of the bilingual program were included in the correlational analysis, which may have led to clustered scores and thus few significant correlations. Second, any test modelled on the MLAT-E/MLAT can be expected to depend quite heavily on literacy, still at early stages of development in 8-year-old children. Finally, it would be possible in principle to claim that aptitude is too unstable in children as young as 8 years to serve as a meaningful predictor – an argument not supported, however, by findings from another study with 8 to 9-year-old children which is described in the following paragraphs.

Tellier and Roehr-Brackin, (2013) worked with an intact group of 8 to 9-year-old children (N = 28) in an English state primary school. The children were randomly assigned to two treatment groups: Group E (N = 14) was taught Esperanto, Group F (N = 14) was taught French. All children were tested for language learning aptitude, metalinguistic awareness and L2 proficiency. The research questions considered whether children would make gains on the measures of aptitude, metalinguistic awareness and L2 proficiency, the relationships between these variables, and any differences between Group E and Group F. Children's language learning aptitude was measured via a slightly modified version of the MLAT-E(UK), the British English version of the MLAT-E (Carroll & Sapon, 2002a). The test of metalinguistic awareness was a short measure comprising two task types: the first required children to pair sentences in different European languages with the same meaning, and the second asked children to translate sentences from three European languages into English. The metalinguistic tasks encouraged the drawing of comparisons between languages, the identification of similarities in form or meaning, and the transfer of knowledge from one language to another. The tests of L2 proficiency were matched tests for the two treatment languages, Esperanto and French. Subtests focused on core vocabulary and structures taught in the L2 sessions and assessed children's skills in reading, writing and listening. The L2 proficiency tests, the test of metalinguistic awareness, and the MLAT-E(UK) were administered at the beginning and end of the school year.

The children progressed on all measures. They made statistically significant gains in L2 proficiency in respectively French or Esperanto, as expected, and in terms of metalinguistic awareness and language learning aptitude. A stepwise multiple regression analysis carried out on the sample as a whole yielded two significant predictors of L2 proficiency at post-test, namely children's performance on the Number Learning subtest of the MLAT-E(UK) at pre-test, which explained 43% of the variance in L2 proficiency, and children's performance on the Matching Words subtest of the MLAT-E(UK) at pre-test, which explained 9%. The Matching Words subtest assesses grammatical sensitivity/language-analytic ability, while the Number Learning subtest assesses memory ability for auditory input. In sum, the study provides evidence that language learning aptitude was dynamic in the young participants, but also that (components of) aptitude can predict L2 achievement. However, the sample size in this study was very small, and further substantiation is clearly needed.

The argument that aptitude is not stable and still developing in young learners has received empirical support elsewhere (Milton & Alexiou, 2006), most notably from a large-scale study (Suárez & Muñoz, 2011) drawing on the MLAT-ES, the Spanish version of the MLAT-E, and the MLAT-EC, the Catalan version. The participants were 629 bilingual Catalan-Spanish children aged 8 to 12. One cohort (N = 325) took the MLAT-ES, the other cohort (N = 304) the MLAT-EC. The test sections in the Spanish and Catalan versions closely resemble the MLAT-E test sections and use the same task types. Results show that children’s scores increase as they mature. A steep and significant increase in scores is evident between ages 8 to 9, especially for the Hidden Words and Matching Words subtests. As children age, the increase in scores becomes less noticeable, i.e. from 10 onwards, gains become smaller. Age 11 may be the starting point of relative stability in aptitude as measured by the MLAT-ES and MLAT-EC – a finding that is broadly in keeping with the research reviewed above (Kiss & Nikolov, 2005; Kiss, 2009).

In addition, a small body of research has considered the specific role of different aptitude components in the development of L2 proficiency (Sáfár & Kormos, 2008; Kormos, 2013). Broadly speaking, it has been argued that younger children rely above all on memory in their L2 learning, while older children and adolescents rely more heavily on language-analytic ability (Harley & Hart, 1997; Robinson, 2005). More recently, Muñoz (2014) reported on the association between young learners’ aptitude and L2 learning outcomes in terms of the four skills of speaking, listening, reading and writing. The researcher worked with bilingual Spanish-Catalan participants (N = 48) who had been learning L2 English from age 6 onwards with about three lessons (or 150 minutes of input) per week. Aptitude was tested by means of the MLAT-ES, since the participant sample was Spanish-dominant. Children were assessed in listening, reading and writing at ages 10-11, and in speaking at ages 11-12. The listening and reading tests had a discrete-item format, while the writing test required pupils to write a short composition. Accuracy, complexity and fluency were scored. The speaking test included a picture description and responding to personal questions.

The results show that overall aptitude scores were correlated moderately but significantly with L2 speaking. Not unexpectedly, the correlation was driven by the Finding Rhymes and Number Learning subtests. Overall aptitude scores correlated more strongly with the other L2 skills, however, and most strongly with L2 writing. In general terms, this finding suggests that the MLAT-ES is a good predictor of general L2 achievement. Analyses by subtest reveal that Number Learning showed the strongest associations, followed by Finding Rhymes and Matching Words, and Hidden Words the weakest ones. Matching Words was most strongly correlated with reading and writing, which suggests that the reading and writing measures allowed learners to employ their language-analytic abilities to the greatest advantage. Overall, aptitude subtests drawing on memory abilities showed only slightly stronger correlations than other aptitude subtests, indicating that 10 to 12-year-old children do rely on memory, but that this is not the only, or even the most critical component; phonetic coding and language-analytic ability were also relevant. In conclusion, the researcher speculates that it may be the case that only children with strong language-analytic abilities will be high achievers (Muñoz, 2014).

In many respects, the construct of language-analytic ability can be linked with the notion of metalinguistic awareness, defined as the ability to focus on and manipulate language form, as well as the ability to treat language as an object of inspection, reflection and analysis (Gombert, 1992; Bialystok, 2001; Baker, 2006). Accordingly, it is worth considering the exact nature of the relationship between the two constructs. Some researchers have simply stated in general terms that aptitude and metalinguistic awareness are partially overlapping concepts (Herdina & Jessner, 2002; Jessner, 2006). Others have focused specifically on language-analytic ability, which, like metalinguistic ability, refers to an individual’s capacity to consider language form in its own right, e.g. by reflecting on language or by reasoning analytically about language (Ranta, 2005; Sawyer & Ranta, 2001). It has been proposed that, if defined in this way, metalinguistic skill and language-analytic ability can be regarded as two sides of the same coin (Ranta, 2002).

Researchers either see aptitude as innate and relatively stable and thus as impacting on the development of metalinguistic awareness (Ranta, 2002), or they assume a bi-directional influence between the two variables, with neither conceptualized as stable

(Jessner, 2006). In addition to language-analytic ability, it has been suggested that the aptitude component of phonetic coding ability may be related to phonological awareness, which, as a component of metalinguistic awareness, is implicated in the development of literacy skills (Sparks & Ganschow, 2001; Kormos, 2013). The theoretically driven argument that (components of) aptitude and metalinguistic awareness are closely related or partially overlapping is consistent with findings suggesting that the role of different aptitude components in L2 learning may change as individuals mature. For instance, the observation that younger children seemingly rely more heavily on memory while older children and adolescents increasingly draw on language-analytic ability would be in keeping with developing metalinguistic abilities and improved literacy skills. However, empirical research on this issue is still in very short supply, in particular with regard to the question of which aptitude component(s) and/or which component(s) of metalinguistic awareness may be associated with the development of specific L2 skills in children under the age of 12.

Justification and objectives of the present study

In summary, empirical evidence to date suggests that aptitude is dynamic in children under the age of 12. As children move through the primary-school years, their scores on aptitude measures gradually increase (Milton & Alexiou, 2006; Kiss, 2009; Suárez & Muñoz, 2011; Tellier & Roehr-Brackin, 2013) with a substantial developmental leap occurring at around ages 8-9. However, with the exception of a single small-scale study (Tellier & Roehr-Brackin, 2013), these findings are exclusively based on cross-sectional research with learners at different ages whose scores were then used to draw inferences about longitudinal developmental patterns.

Furthermore, existing evidence suggests that different aptitude components may be differentially associated with success on different dimensions of L2 proficiency (Sáfár &

Kormos, 2008; Kormos, 2013; Muñoz, 2014), reflecting findings from adult populations. Broadly speaking, it has been proposed that younger children rely more on memory in their L2 learning, while older children and adolescents favor language-analytic ability (Harley & Hart, 1997; Robinson, 2005). Empirical evidence for this theoretically plausible argument is still rather limited, though. Finally, while theoretical arguments about the relationship and potential overlap between language-analytic ability and metalinguistic awareness have been put forward (Ranta, 2005; Sawyer & Ranta, 2001), there is still only little empirical evidence about the association of these two abilities in child L2 learners, with only a single small-scale study (Tellier & Roehr-Brackin, 2013) operationalizing and measuring both constructs in the context of the same research project.

With this in mind, we sought to address from an empirical perspective (1) the conceptual question of the relationship between aptitude and metalinguistic awareness by measuring both constructs within the same sample of children; (2) the issue of stability or otherwise of aptitude by directly assessing development over time within the same sample of children; and (3) the question of whether and to what extent aptitude can predict L2 achievement in young L2 learners, and in particular which aptitude components would be associated with achievement in which L2 skills. Specifically, we posed the following research questions:

(1) What is the relationship between language learning aptitude and metalinguistic awareness in 8 to 9-year-old primary-school children?

(2) Are children's levels of language learning aptitude and metalinguistic awareness stable over time?

(3) Does language learning aptitude predict children's achievement in L2 French? If it does, which aptitude components are associated with which L2 skills?

Methodology

The research questions were addressed in the context of a year-long quasi-experimental study with primary-level L2 learners.

Participants and treatment

The study was carried out with intact classes of 8 to 9-year-old monolingual English-speaking children (N = 111) from five different English primary schools. Nearly 90% of primary schools in England provide fewer than 60 minutes of L2 instruction per week, with as little as 30 minutes per week not at all uncommon (Board & Tinsley, 2017). The participating schools agreed to accommodate 75 minutes of language input per week for the duration of the present study, divided into a weekly 60-minute language lesson taught by a teacher specifically employed for the project and 15 minutes of follow-up work led by the usual class teachers and scheduled at a time that was convenient for them. For the purpose of another part of the project, the children were divided into four groups and consecutively taught two L2s over the school year. In Phase 1, the four groups were taught, respectively, German, Italian, Esperanto, or Esperanto with a focus-on-form element. In Phase 2, all groups were taught French with a focus-on-form element. Each instructional phase lasted 16 weeks and comprised 20 hours of teaching in total.

In classes with a focus-on-form element, i.e. one of the Esperanto classes in Phase 1 and all French classes in Phase 2, approximately 40 minutes of the weekly lesson were devoted to the same program content as in the other three language groups, while approximately 20 minutes were spent on a focus-on-form activity drawing on the same topic area, but concerned with formal characteristics of the target language (Esperanto or French), e.g. adjectival inflection, case marking, word order, or grammatical gender. Children were scaffolded through the noticing and solving of language-related problems, e.g. identifying a rule and applying it, constructing new words from stems and affixes, or appreciating semantic relationships between words. Aside from the presence or absence of form-focused activities, the instructional activities were the same across all languages.

All instructional materials were tailor-made for the project to ensure comparability across languages in terms of targeted vocabulary and structures as well as content-based progression in the course of the school year. The L2 lessons covered all four skills (Cameron, 2001), with a story-based instructional approach (Adair-Hauck & Donato, 2009). The lessons were designed to be incremental to maximize children's opportunities to create links with previous knowledge, and the story-based content allowed children to not only acquire lexical items, but also exposed them to the morphology and syntax of the language and a certain amount of connected discourse.

The language materials were designed to cover more than one curriculum objective to ensure that maximum use was made of the curriculum time allocated to the study. The language program in Phase 1 focused on the topic of natural habitats, including the characteristics of animals, nutrition, life cycle, and simple classification and was linked so curriculum objectives in science; the language program in Phase 2 focused on travel, clothing, weather, and locations and was linked to curriculum objectives in geography.

Teaching and learning activities included the completion of task sheets and a project book; children worked with reference cards, engaged in games, role-plays, filming, and interviews. Songs were used to practice idioms, provide variety and encourage motivation. Activities were carried out in the target language, while instructions were typically given in English to ensure easy comprehension and the allocation of as much time as possible to the L2 program content. Sample instructional materials can be found in the IRIS digital repository of research instruments at https://www.iris-database.org/.

The focus in the present paper is on the role of aptitude and metalinguistic awareness in the children's learning of L2 French, i.e. the language all participants were exposed to in Phase 2 of the project.

Instruments and procedures

The children were tested for language learning aptitude and metalinguistic awareness at the beginning and end of Phase 1, henceforth respectively referred to as pre-test and post-test for convenience. In addition, the children's French proficiency was assessed at the beginning and end of Phase 2. We administered a pre-test, an immediate post-test and a delayed post-test 8 weeks later. All testing took place during regular class time. The measures that were used are described in detail below.

Language learning aptitude was measured by means of a slightly amended version of the MLAT-E(UK) designed for speakers of British English, which is a validated version of the original American English MLAT-E (Carroll & Sapon, 2002a). The test is a paper-and-pen measure comprising four subtests with a total maximum score of 130. The *Number Learning* subtest (max. score 25) assesses phonetic coding ability and memory ability for sound-meaning relationships. The test comprises 25 items to be completed within four minutes. Children are taught the names for numbers in an invented language immediately before the test. For the test itself, children record in figures numbers spoken aloud by the administrator. The *Finding Rhymes* subtest (max. score 45) measures the ability to perceive and distinguish between English speech sounds. The test presents 45 items to be completed within six minutes. Children must choose one of four possible alternatives which they consider to be the best rhyme for each keyword. The *Hidden Words* subtest (max. score 30) measures the ability to associate sounds with symbols as well as knowledge of English vocabulary. The test, to be completed within five minutes, presents 30 keywords which are written approximately as pronounced. Children must choose which of four correctly spelled words matches each ‘hidden word’ most closely in meaning. Finally, the *Matching Words* subtest (max. score 30) measures grammatical sensitivity and language-analytic ability. The test presents 30 items to be completed within 18 minutes. Children’s attention is drawn to a keyword in an English sentence. They are asked to choose a ‘matching word’ from a second sentence presented underneath. The MLAT-E took approximately 50 minutes to administer and showed excellent reliability: Cronbach’s α = .960 (pre-test) and .962 (post-test).

Metalinguistic awareness was assessed by means of a dedicated paper-and-pen measure developed for English-speaking children aged 8 to 11 that comprises eleven tasks. In the present study, some groups did not finish Task 11, so all analyses are based on Tasks 1-10, which together yielded a maximum possible score of 83. The test focuses on knowledge about domains relevant to both L1 and L2 learning, such as lexical semantics, morphology, syntax, linguistic ambiguity, and basic metalinguistic terminology. It targets concepts relevant to the L2 learning of English-speaking children, who are typically exposed to European languages, e.g. grammatical gender, case, verbal and adjectival agreement, word order, cognates, and similarities and differences between languages. The tasks draw on different European languages and a language specifically constructed for the test. The measure was piloted extensively and has been found to exhibit strong reliability and validity (Tellier, 2013). The test is available from the IRIS digital repository of research instruments at https://www.iris-database.org/. In the present study, the test took about 60 minutes to administer, including instructions, explanations and examples. It showed very good reliability: Cronbach’s α = .831 (pre-test) and .995 (post-test).

The children’s *L2 French proficiency* was assessed by means of a paper-and-pen measure developed for the study in accordance with recommendations for age-appropriate language testing (Hasselgren & Caudwell, 2016). The test was aligned to recognized international and national descriptors of language competence. While the Common European

Framework of Reference (CEFR; Council of Europe, 2004) is the most widely-recognized descriptive framework in the European context, it was not designed specifically with primary-level learners in mind and was therefore only partly suitable as a reference point (Hasselgren & Caudwell, 2016). Therefore, we additionally drew on the Asset Languages Breakthrough stage descriptors (Asset, 2005), which map onto sub-divisions of CEFR Level A1 and are thus more in line with young children's expected progress and achievements in a limited-input setting.

The French test comprised four sections aimed at assessing listening (max. score 12), reading (max. score 16), writing (max. score 34), and grammar (max. score 20) and yielded a total maximum score of 82. In summary, the listening and grammar sections were based on multiple-choice items, taking into account the participating children's emerging literacy skills; the reading section included multiple-choice items and simple reading comprehension questions in L1; the writing section required accurate copying and simple sentence construction based on given key words, again taking into account children's developing literacy skills. All task types were chosen with children's as yet very low proficiency in French in mind, while at the same time giving scope for maximum discrimination. The test is available from the IRIS digital repository of research instruments at https://www.iris-database.org/. The construction of each test section is described in detail in what follows.

The listening section was graded in difficulty. Earlier items presented simple concrete nouns together with the appropriate indefinite article (e.g. *une carte*); later items included numbers, plural nouns and verbs, prepositional phrases and/or adjectives (e.g. *Il y a de la neige dans la forêt*). Each item was read aloud twice by the test administrator. Children were required to circle whichever response picture they felt most closely matched the auditory cue. Each child was provided with a card strip to cover the picture responses of other items, if desired, to help orientation and focus.

The reading section comprised four tasks of which two were in multiple-choice format and based on short, simple sentences and formulaic expressions. Children were asked to circle one of four pictures which best matched the written text. A further task type required children to read four short sentences (*Je suis petit et brun. Je mange des vers. Aujourd'hui il neige. Je dors sous les feuilles.*) and to answer four questions based on the reading (*What am I like? What do I eat? What is the weather like? Where do I sleep?*). Children were instructed to write their responses in English in order to achieve an exclusive focus on reading ability in the L2 (rather than reading and writing in combination). The final reading task consisted of a short story: *Au Canada, beaucoup d'animaux habitent dans la forêt. Il y a beaucoup de sapins et de pins. Le lac est très grand. Un castor nage dans la rivière bleue.* Children were asked to record in English two points or facts gathered from their reading.

The writing section presented three tasks. In the first task, children were required to accurately copy a compound noun (*un porc-épic*). To gain maximum points, a child had to correctly place all script features such as diacritics and punctuation, and the word had to be comprehensibly legible. A picture provided context. The remaining two tasks presented four key words in French alongside English scaffolding. An example showed two sentences about an odd-looking animal. The test administrator talked the children through the example, showing how the stimulus words and additional words had been combined to construct an interesting response. Children were encouraged to use the example sentences for support and to use additional lexical items from their own repertoire to construct further 'interesting' sentences about the odd-looking animal. This gave children of lower ability support to write very simple sentences (e.g. subject – verb) while allowing children of higher ability more scope and encouragement to expand and produce sentences containing adjectives and prepositional phrases, for instance. The scoring reflected this flexibility, with more 'interesting' sentences awarded more points.

The grammar section was in multiple-choice format throughout and required children to choose one of three possible response options for each item. Color pictures and example items, where appropriate, provided scaffolding. Careful construction of each item ensured that the context conveyed by the item clearly indicated the use of a particular form. The word *deux*, for instance, with a number 2 in brackets afterwards, clearly indicated the need to construct the plural form of a noun. The test items focused on three grammatical features covered in the language program: plurals of nouns, verb conjugation, and grammatical gender. The test administrator read aloud the stimuli, but not the responses. Auditory input was considered necessary to ensure that children's knowledge of grammar was tested as independently as possible from their reading and writing skills.

Overall, the test took about 60 minutes to administer. It was piloted prior to use and showed excellent reliability in the present study: Cronbach's α = 1.000 (pre-test), .999 (immediate post-test) and 1.000 (delayed post-test).

Data analysis

One-sample Kolmogorov-Smirnov (KS) tests showed that data sets arising from the measures of language learning aptitude, metalinguistic awareness and L2 French proficiency did not significantly diverge from a normal distribution, so parametric statistical analyses were used. However, several data sets based on subtest scores yielded significant KS test results, so we used non-parametric tests in analyses involving these data sets.

In order to address RQ1 (What is the relationship between language learning aptitude and metalinguistic awareness in 8 to 9-year-old primary-school children?), we first present descriptive statistics for the relevant variables. We then ran bivariate correlations and conducted a principal components analysis on the MLA test (pre-test) and MLAT-E subtest scores (pre-test). The assumptions of the procedure were met, as follows: Kaiser-Meyer-Olkin measure of sampling adequacy = .782; Bartlett's test of sphericity $p < .001$.

In order to address RQ2 (Are children's levels of language learning aptitude and metalinguistic awareness stable over time?), we first present descriptive statistics for the relevant variables. We then ran paired-samples t-tests for normally distributed data sets and Wilcoxon Signed Rank tests for data sets diverging from a normal distribution to compare scores across testing times. We additionally ran bivariate correlations between pre-test and post-test scores on each measure to establish the extent to which children's levels of performance at different testing times were associated.

In order to address RQ3 (Does language learning aptitude predict children's achievement in L2 French? If it does, which aptitude components are associated with which L2 skills?), we first present descriptive statistics and then ran paired-samples t-tests for normally distributed data sets and Wilcoxon Signed Rank tests for data sets diverging from a normal distribution to compare performance in French across testing times. Subsequently, we report bivariate correlations between the variables under study. Finally, we conducted a linear regression analysis (enter method) with the MLAT-E subtest scores as the independent variables and French gain scores as the dependent variable. The assumptions of a normal distribution of the data, homogeneity of variances, linearity, and the absence of multicollinearity were checked following the recommendations in Larson-Hall (2016) and found to be met to a reasonable standard.

For all analyses, the alpha level was set at .05. Effect sizes are reported as Cohen's d and interpreted in accordance with field-specific benchmarks (Plonsky & Oswald, 2014).

Results

The first research question asked about the relationship between language learning aptitude and metalinguistic awareness. The descriptive statistics for children's performance on the tests of aptitude (MLAT-E) and metalinguistic awareness (MLA), are shown in Table 1 and Table 2, respectively.

[TABLE 1 NEAR HERE]

[TABLE 2 NEAR HERE]

The relatively low facility values at pre-test, i.e. prior to any L2 instruction, indicate that both tests were quite challenging for the young learners. At the same time, there is clear variation in performance, especially on the aptitude test, suggesting individual differences in the sample. In order to answer RQ1, children's scores on the MLAT-E and the MLA test were correlated. The results show a significant relationship of medium strength at pre-test, $r = .524^{**}$, $p < .001$, and a strong significant relationship at post-test, $r = .703^{**}$, $p < .001$. The increase in correlational strength between the beginning and end of Phase 1 of the study suggests that children's gradually increasing cognitive maturity, their L2 learning experience, and/or both in combination seem to have led to a closer association between the abilities measured.

Correlations between the four individual MLAT-E subtests (pre-test) and the MLA test as a whole (pre-test) were significant at the .01 level throughout and ranged from .33 to .44 in strength, indicating that all aptitude components share some common variance with metalinguistic awareness. The fact that the subtest correlations are similar in strength may be counter-intuitive at first glance, since the Matching Words subtest might be expected to correlate more strongly with metalinguistic awareness in view of its focus on language-analytic ability. However, all MLAT-E subtests except the Number Learning subtest rely on literacy skills to some extent, as does the MLA test. Given the young age of the participants and the fact that their reading and writing skills were still developing, the relatively heavy reliance on literacy skills of most of the measures used may have been a more important factor than language-analytic vs. memory and/or phonetic abilities. [1]

In a second step towards addressing RQ1, we conducted a principal components analysis. The resulting model yielded a single component with an eigenvalue > 1 and a clear break after that component visible in the associated scree plot. The extracted component (eigenvalue = 2.807) explains 56% of the variance, with factor loadings ranging from .692 to .861. Thus, the MLAT-E subtests and the MLA test loaded onto the same component, which is indicative of considerable common variance. Put differently, the constructs of language learning aptitude and metalinguistic awareness as operationalized in the present study are indeed overlapping.

The second research question asked whether children's levels of language learning aptitude and metalinguistic awareness would remain stable over time. The descriptive statistics (see Table 1) suggest an improvement in aptitude test performance on the second administration, which is confirmed by a paired-samples t-test yielding a significant medium effect, $t(110) = -13.378$, $p < .001$, Cohen's $d = .691$. Unsurprisingly, the children also improved significantly on each of the MLAT-E subtests, Hidden Words: $t(110) = -12.497$, $p < .001$, Cohen's $d = .971$; Number Learning, Finding Rhymes and Matching Words: Wilcoxon Signed Rank tests all $p < .001$.

While the range and standard deviation of the MLAT-E did not change much over time, the range and standard deviation of the test of metalinguistic awareness clearly increased (see Table 2), indicating greater heterogeneity among the children at post-test. As in the case of the MLAT-E, mean test performance improved significantly on the second administration, as confirmed by a paired-samples t-test showing a significant medium effect, $t(110) = -8.637$, $p < .001$, Cohen's $d = .677$. Thus, we can conclude that both language learning aptitude and metalinguistic awareness are dynamic in the young participants, with statistical progress in evidence over Phase 1 of the study.

We additionally correlated pre- and post-test scores on each of the measures. MLAT-E scores were very strongly and significantly associated at pre-test and post-test, $r = .853^{**}$, $p < .001$, and MLA test scores were likewise significantly associated at pre-test and post-test, also yielding quite a strong coefficient, $r = .684^{**}$, $p < .001$. These results suggest that children scoring highly on a measure at pre-test also scored highly at post-test, and, conversely, that children who performed poorly tended to do so on both occasions. This is particularly true for the MLAT-E, but also applies to the MLA test, though to a slightly lesser extent. This pattern of findings seems to suggest that despite the dynamicity of aptitude in our sample, a measure of the construct may still serve as a useful predictor. This leads us to the third and final research question, which asked whether language learning aptitude would predict children's achievements in L2 French.

As children's performance in French proficiency at immediate and delayed post-test did not differ statistically, $t(110) = .205$, $p = .838$, and as the two sets of scores were strongly correlated, $r = .837^{**}$, $p < .001$, we calculated a mean from both tests to give a combined post-test score. Gain scores were then calculated by deducting pre-test scores from combined post-test scores. In cases where a child had completed only one of the post-tests, that score was used as their combined post-test score. Table 3 shows the descriptive statistics for children's performance on the L2 French proficiency test.

[TABLE 3 NEAR HERE]

The results in Table 3 indicate that the French test was very challenging at pre-test – not unexpectedly, since it was administered when French was first introduced and when children had had no or extremely limited prior exposure, e.g. through lessons in the previous school year or participation in language clubs. It is worth noting that there is considerable variation in performance both at pre-test and at post-test, with some children performing very poorly and others doing very well indeed. It is also interesting to note the presence of negative gains, i.e. some children did worse at post-test than at pre-test. This pattern of results was also in evidence in the descriptive statistics for all the French subtests, although it is most pronounced for writing and grammar, less so for the receptive skills of listening and reading. The mean gain score for French proficiency overall indicates an improvement over the course of Phase 2, as expected. A paired-samples t-test confirms that this improvement is large and statistically significant, $t(110) = -19.068$, $p < .001$, Cohen's $d = 1.673$. Significant gains are likewise in evidence for all subtests, Wilcoxon Signed Rank tests $p < .001$ in all cases.

In order to address our research question about the predictive power or otherwise of language learning aptitude, we ran a correlation between the children's MLAT-E scores (post-test, i.e. at the start of Phase 2) and their French proficiency gain scores (with French instruction taking place during Phase 2). The result shows a significant association of medium strength, $r = .531^{**}$, $p < .001$, which indicates that aptitude can indeed predict L2 achievement in our young learners. In order to establish which components of aptitude might predict which specific L2 skills, we ran further correlations by subtest, as shown in Table 4.

[TABLE 4 NEAR HERE]

The results in Table 4 show that the overall aptitude score significantly correlates with gains in French reading, grammar and listening, but not writing. This pattern is reflected in the correlations based on the aptitude subtests, which show statistical associations with gains in grammar, reading and listening at moderate to medium levels of strength, but not with gains in writing. Matching Words, the measure of grammatical sensitivity/language-analytic ability, correlates more strongly with grammar and reading than with listening, which is unsurprising. The same pattern can be observed for the other three sub-tests, however, which is perhaps more surprising, at least at first glance. Number Learning, Finding Rhymes and Hidden Words consistently correlate more strongly with grammar and reading than with listening, despite their phonetic components. However, as pointed out above, all MLAT-E subtests except Number Learning rely to some extent on literacy, a skill that is still developing in 8 to 9-year-olds and that is clearly reliant on phonological awareness. Overall, Matching Words and Finding Rhymes yield slightly stronger coefficients than Number Learning and Hidden Words.

As a final step towards addressing our research question, we conducted a linear regression analysis (enter method) with the MLAT-E subtests as the independent variables and French gains as the dependent variable. The resulting model ($r = .56$, $R^2 = .313$) explains 31% of the variance in L2 French gains and is summarized in Table 5.

[TABLE 5 NEAR HERE]

The results in Table 5 indicate that two aptitude subtests contribute significantly to the model, with Matching Words uniquely explaining 6% of the variance and Finding Rhymes uniquely explaining 4% of the variance in French gain scores. Number Learning approaches significance, explaining 2% of the variance. [2] In summary, the results of the regression analysis confirm that our measure of language learning aptitude significantly predicts the young participants' progress in L2 French, explaining a total of about a third of the variance in children's gain scores. With regard to specific aptitude components, we found that the MLAT-E subtests of grammatical sensitivity/language-analytic ability and phonetic coding ability/phonological awareness made unique significant contributions to explaining the variance in scores.

Discussion and conclusions

It is now possible to draw together the results in order to arrive at conclusions about the key research issues addressed in the present study. We begin with the question of stability or otherwise of aptitude and metalinguistic awareness in young learners and then consider to what extent aptitude measures may serve as predictors of L2 achievement in instructed young learners.

The dynamicity of aptitude and metalinguistic awareness in instructed child learners

The first finding of note is empirical support for the argument that language learning aptitude and metalinguistic awareness are overlapping constructs in young learners (Ranta, 2005; Sawyer & Ranta, 2001) when they are measured, respectively, by means of the MLAT-E(UK) and a test of metalinguistic awareness assessing a range of metalinguistic skills. In line with arguments made previously, we agree that language-analytic ability is a key source of overlap. In addition, we have suggested that in the case of our 8 to 9-year-old participants at least, the fact that both measures draw on children's still developing literacy skills may additionally help explain the common variance. Language-analytic ability is strongly implicated in the test of metalinguistic awareness we used as well as the Matching Words subtest of the MLAT-E, while phonological awareness is strongly implicated in the Hidden Words and Finding Rhymes subtests of the MLAT-E as well as in the ability to learn to read and write alphabetic languages. Memory ability appears to be a fundamental capacity that is drawn upon in the MLAT-E, the test of metalinguistic awareness and the development of literacy skills to the extent that children must be able to acquire and retrieve constructions of the English language, whether at word level, at sub-lexical level or in terms of larger chunks extending beyond single words.

Our second finding of note is that both language learning aptitude and metalinguistic awareness are dynamic in young learners aged 8-9 years, with children improving significantly on the second administration of each test. [3] Unlike previous research which has reported this conclusion (Milton & Alexiou, 2006; Kiss, 2009; Suárez & Muñoz, 2011), our findings are based on data obtained from the same sample of children. We therefore have direct longitudinal evidence from a reasonably large sample of participants (N = 111) that does not rely on inferences drawn from cross-sectional data. Our evidence is consistent with the findings of the only other study we are aware of which likewise obtained longitudinal data (Tellier & Roehr-Brackin, 2013), though from a much smaller sample of children (N = 28).

Taken together, the findings that (1) language learning aptitude and metalinguistic awareness show sufficient common variance to load onto the same factor in a principal components analysis, and (2) language learning aptitude and metalinguistic awareness are dynamic in young learners do not sit well with Ranta's (2002) argument that language learning aptitude may be an innate and relatively stable trait that affects the development of metalinguistic awareness. Although we found a correlation between the two variables, not unexpected on the basis of this argument, we also found that children showed significant improvements on both measures. Thus, it may be more appropriate to conclude that the two variables interact in a cyclical manner, i.e. we are probably and more realistically dealing with a relationship of mutual influence in the sense of connected growth, e.g. as conceptualized in dynamic systems/complexity theory (de Bot, Lowie, & Verspoor, 2007; Larsen-Freeman, 2009), rather than with a one-way cause-effect relationship. What remains unresolved for now is whether these improvements are due to children's L2 learning experience, to general cognitive development over time regardless of any L2 learning, or to both in combination. Theoretical considerations would lead us to expect that L2 learning experience might be an important – though probably not the only – factor here, especially in the context of the development of metalinguistic awareness (Herdina & Jessner, 2002; Jessner, 2006), while the simultaneous development of literacy skills can also be expected to play a decisive role.

The predictive power of aptitude in instructed child learners

Even though we report evidence for the dynamicity of language learning aptitude and metalinguistic awareness in young learners, we also uncovered strong correlations between pre-test and post-test performance on measures of these two constructs, which suggest that the MLAT-E, for instance, may serve as a useful predictor. Accordingly, our third finding of note is a statistical correlation between children's MLAT-E test scores at the start of L2 instruction and their subsequent gains in L2 French. This result confirms that aptitude can indeed predict L2 achievement (Kiss & Nikolov, 2005; Muñoz, 2014) even in child learners as young as 8-9 years. However, as already indicated, we must bear in mind the theoretical complexities of the interplay we are observing. We have established that aptitude itself is not (yet?) stable. Moreover, we cannot exclude the possibility that general cognitive development taking place over time, developing literacy skills and experience with L2 learning have a washback effect, so it is important to emphasize that we do not claim a simple unilateral cause-effect relationship between aptitude and L2 gains. As long as this caveat is taken into consideration, however, it is certainly possible to make use of an aptitude measure in order to predict likely classroom achievement if this is desirable for practical purposes in a particular educational context, for instance.

Our fourth finding of note is that children's overall MLAT-E scores as well as all MLAT-E subtest scores correlated significantly with gains in French grammar, reading and listening, but not writing – a pattern of results in keeping with findings from a meta-analysis of aptitude research in adolescents and adults (Li, 2016) where L2 writing was not correlated with total aptitude scores, although the other three skills did correlate, as did grammar and vocabulary knowledge. Moreover, our result contradicts findings reported in Muñoz (2014), a study otherwise comparable to ours which used the MLAT-ES and involved instructed child learners, though at ages 10-12 they were older than ours. Muñoz (2014) reports the strongest correlation between aptitude overall with writing in L2 English, compared with the other three skills, though it is worth noting that writing was measured in terms of complexity, accuracy and fluency in that study. [4] In terms of the overall variance explained, our regression model yielded a slightly stronger coefficient at r = .56 than the coefficient of r = .49 reported in Li's (2016) meta-analysis. Our overall result is almost the exact mean of the weakest and strongest correlations reported in Muñoz' (2014) study, where coefficients ranged from r = .36 for speaking to r = .75 for writing.

Our final finding of note is that, taken together, the MLAT-E subtests explained nearly a third of the variance in children's overall L2 French proficiency gains. In itself, this finding is relatively unsurprising, given the results of our correlation analysis reported above. Essentially, it further confirms the conclusion that aptitude can indeed predict L2 achievement, even in young learners (Kiss & Nikolov, 2005; Muñoz, 2014). More interestingly perhaps, and certainly more surprisingly, two MLAT-E subtests, Finding Rhymes (explaining 4% of the variance) and Matching Words (explaining 6% of the variance), uniquely contributed to the regression model; Number Learning only approached significance. This pattern of results is very different from the findings reported in the two most directly comparable studies. In both Muñoz (2014) and Tellier and Roehr-Brackin (2013), Number Learning showed the strongest association with L2 achievement, pointing towards a predictive role for memory ability in the L2 achievement of instructed child learners, whereas a measure of language-analytic ability emerged as the strongest predictor in the present study, followed by a measure of phonetic ability/phonological awareness.

Overall, then, the findings pertaining to the predictive power of aptitude in relation to L2 gains achieved by our 8 to 9-year-old participants more closely match an ‘adult’ pattern than a ‘child’ pattern: There is no correlation between overall aptitude and L2 writing, and the most important aptitude component for predicting overall L2 gains is language-analytic ability. Clearly, this is not consistent with the argument that young learners with their reliance on predominantly implicit learning mechanisms primarily draw on memory ability, with language-analytic abilities only becoming more important from around age 12 onwards when explicit learning mechanisms kick in fully. It is, however, consistent with another recent line of argument which suggests that the type of instruction young learners are exposed to may be just as important as chronological age in determining the use of primarily implicit vs. primarily explicit learning mechanisms (Lichtman, 2013, 2016). In other words, the more form-focused the instruction learners experience, the more relevant language-analytic ability becomes, and the greater its role in children’s L2 success.

Implications for further research

While the present study has revealed insights into the role of language-analytic ability in child L2 classroom learning, some questions remain unanswered. Specifically, future research should include different instructional conditions in the same design with aptitude and metalinguistic awareness measures, so results can shed light onto the interaction between implicit vs. explicit learning and the relative importance of memory vs. language-analytic abilities in young learners. Ages 8-9 would appear to be a suitable age group, given the results of the present study as well as the previous finding that this may be an important age range in the development of language-analytic ability (Suárez & Muñoz, 2011). This age range also coincides with the completion of a shift from pre-operational to concrete-operational thought (Piaget, 1929; Gruber & Vonèche, 1977; Anderson, 2005) and the development of reading and writing skills. Measures of language learning aptitude could potentially be extended to include tests of implicit learning ability (Wen, Biedroń, & Skehan, 2017), such as serial reaction time tasks (Granena, 2013, 2014) or online learning tasks using linguistic stimuli, e.g. the LLAMA D subtest of sound recognition (Meara, 2005), if piloting shows that children as young as 8-9 can cope with these measures. Finally, including a control group that does not receive L2 instruction in a research design that incorporates measures of aptitude and metalinguistic awareness would help identify to what extent the development of these two abilities over time is dependent on L2 experience.

Notes

[1] There were no significant correlations between children's MLAT-E and MLA test scores and the most recent literacy scores available to us, that is, school-internal reading and writing scores obtained two years prior to the administration of the MLAT-E and the MLA test.

[2] The correlation coefficients squared in the last column of Table 5 show the unique contribution of each independent variable, while R^2 includes shared variance between the independent variables. Therefore, the unique variances do not add up to the total variance explained; instead, they show the relative importance of each predictor.

[3] A reviewer rightly pointed out that the increase in scores may be due to a test-retest effect. Unfortunately, our research design does not allow us to disentangle a possible test-retest effect from the maturation account we favor. As several months went by between testing times, the young participants will have developed in terms of both their cognitive and their linguistic abilities.

[4] A reviewer noted that the basic nature of the writing subtest used in the present study, which reflects the relatively low proficiency level of the participating children, might be a possible reason for the absence of a significant correlation.

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.

Adair-Hauck, B., & Donato, R. (2009). Using a story-based approach to teach grammar. In J. L. Shrum & E. W. Glisan (Eds.), *Teachers' handbook: Contextualised foreign language instruction K-12* (4th ed., pp. 216-244). Boston, MA: Heinle & Heinle.

Anderson, J. R. (2005). *Cognitive psychology and its implications* (6th ed.). New York, NY: Worth Publishers.

Asset (2005). *Breakthrough French primary teacher assessment materials*. Cambridge: Oxford, Cambridge and RSA Examinations Ltd.

Baker, C. (2006). *Foundations of bilingual education and bilingualism* (4th ed.). Clevedon: Mutilingual Matters.

Bialystok, E. (2001). *Bilingualism in development: Language, literacy, and cognition*. Cambridge: Cambridge University Press.

Board, K., & Tinsley, T. (2017). *Language trends 2016/17: Language teaching in primary and secondary schools in England.* Manchester: British Council.

Cameron, L. (2001). *Teaching language to young learners*. Cambridge: Cambridge University Press.

Carroll, J. B. (1962). The prediction of success in intensive foreign language training. In R. Glaser (Ed.), T*raining research and education* (pp. 87-136). New York, NY: John Wiley & Sons.

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). Rowley, MA: Newbury House.

Carroll, J. B. (1990). Cognitive abilities in foreign language aptitude: Then and now. In T. S. Parry & C. W. Stansfield (Eds.), *Language aptitude reconsidered* (pp. 11-29). Englewood Cliffs, NJ: Prentice Hall.

Carroll, J. B., & Sapon, S. (1959). The Modern Language Aptitude Test. San Antonio, TX: Psychological Corporation.

Carroll, J. B., & Sapon, S. M. (2002a). Modern Language Aptitude Test - Elementary: MLAT-E. Rockville, MD: Second Language Testing Foundation.

Carroll, J. B., & Sapon, S. M. (2002b). Modern Language Aptitude Test: MLAT. N. Bethesda, MD: Second Language Testing Inc.

Council of Europe (2004). *Common European Framework of Reference for Languages: Learning, teaching, assessment*. Cambridge: Cambridge University Press.

de Bot, K., Lowie, W., & Verspoor, M. (2007). A dynamic systems theory approach to second language acquisition. *Bilingualism: Language and Cognition, 10(1)*, 7-21.

DeKeyser, R. M. (2000). The robustness of critical period effects in second language acquisition. *Studies in Second Language Acquisition, 22*, 499-533.

Dörnyei, Z., & Skehan, P. (2003). Individual differences in second language learning. In C. J. Doughty & M. H. Long (Eds.), *The handbook of second language acquisition* (pp. 589-630). Malden, MA: Blackwell.

García Mayo, M. d. P. (2003). Age, length of exposure and grammaticality judgements in the acquisition of English as a foreign language. In M. d. P. García Mayo & M. L. García

Lecumberri (Eds.), *Age and the acquisition of English as a foreign language* (pp. 94-114). Clevedon: Multilingual Matters.

Gombert, J. E. (1992). *Metalinguistic development*. Hemel Hempstead: Harvester.

Granena, G. (2013). Individual differences in sequence learning ability and second language acquisition in early childhood and adulthood. *Language Learning, 63(4)*, 665-703.

Granena, G. (2014). Language aptitude and long-term achievement in early childhood L2 learners. *Applied Linguistics, 35(4)*, 483-503.

Gruber, H. E., & Vonèche, J. J. (Eds.). (1977). *The essential Piaget*. London: Routledge.

Harley, B., & Hart, D. (1997). Language aptitude and second language proficiency in classroom learners of different starting ages. *Studies in Second Language Acquisition, 19*, 379-400.

Hasselgren, A., & Caudwell, G. (2016). *Assessing the language of young learners*. Sheffield: Equinox.

Herdina, P., & Jessner, U. (2002). *A dynamic model of multilingualism: Perspectives of change in psycholinguistics*. Clevedon: Multilingual Matters.

Jaekel, N., Schurig, M., Florian, M., & Ritter, M. (2017). From early starters to late finishers? A longitudinal study of early foreign language learning in school. *Language Learning, 67(3)*, 631-664.

Jessner, U. (2006). *Linguistic awareness in multilinguals: English as a third language*. Edinburgh: Edinburgh University Press.

Johnson, J. S., & Newport, E. L. (1989). Critical period effects in second language learning: The influence of maturational state on the acquisition of English as a second language. *Cognitive Psychology, 21*, 60-99.

Kiss, C. (2009). The role of aptitude in young learners' foreign language learning. In M. Nikolov (Ed.), *The age factor and early language learning* (pp. 253-276). Berlin: Mouton de Gruyter.

Kiss, C., & Nikolov, M. (2005). Developing, piloting, and validating an instrument to measure young learners' aptitude. *Language Learning, 55(1)*, 99-150.

Kormos, J. (2013). New coneptualizations of language aptitude in second language attainment. In G. Granena & M. H. Long (Eds.), *Sensitive periods, language aptitude, and ultimate L2 attainment* (pp. 131-152). Amsterdam: John Benjamins.

Larsen-Freeman, D. (2009). Adjusting expectations: The study of complexity, accuracy, and fluency in second language acquisition. *Applied Linguistics, 30(4)*, 579-589.

Larson-Hall, J. (2008). Weighing the benefits of studying a foreign language at a younger starting age in a minimal input situation. *Second Language Research, 24(1)*, 35-63.

Larson-Hall, J. (2016). *A guide to doing statistics in second language research using SPSS and R.* New York, NY: Routledge.

Li, S. (2016). The construct validity of language aptitude: A meta-analysis. *Studies in Second Language Acquisition, 38(4)*, 801-842.

Lichtman, K. (2013). Developmental comparisons of implicit and explicit language learning. *Language Acquisition, 20(2)*, 93-108.

Lichtman, K. (2016). Age and learning environment: Are children implicit second language learners? *Journal of Child Language, 43*, 707-730.

Meara, P. (2005). Llama Language Aptitude Tests. University of Wales Swansea.

Milton, J., & Alexiou, T. (2006). Language aptitude development in young learners. In C. Abello-Contesse, R. Chacón-Beltrán, M. D. López-Chiménez, & M. M. Torreblanca-López (Eds.), *Age in L2 acquisition and teaching* (pp. 177-192). Oxford: Peter Lang.

Muñoz, C. (2006). The effects of age on foreign language learning: The BAF project. In C. Muñoz (Ed.), *Age and the rate of foreign language learning* (pp. 1-40). Clevedon: Multilingual Matters.

Muñoz, C. (2008). Age-related differences in foreign language learning: Revisiting the empirical evidence. *IRAL, 46*, 197-220.

Muñoz, C. (2009). Input and long-term effects of early learning in a formal setting. In M. Nikolov (Ed.), *The age factor and early language learning* (pp. 141-159). Berlin: Mouton de Gruyter.

Muñoz, C. (2014). The association between aptitude components and language skills in young learners. In M. Pawlak & L. Aronin (Eds.), *Essential topics in applied linguistics and multilingualism: Studies in honour of David Singleton* (pp. 51-68). Cham: Springer.

Piaget, J. (1929). *The child's conception of the world*. London: Routledge.

Plonsky, L., & Oswald, F. L. (2014). How big is 'big'? Interpreting effect sizes in L2 research. *Language Learning, 64*(4), 878-912.

Qureshi, M. A. (2016). A meta-analysis: Age and second language grammar acquisition. *System, 60*, 147-160.

Ranta, L. (2002). The role of learners' language analytic ability in the communicative classroom. In P. Robinson (Ed.), *Individual differences and instructed language learning* (pp. 159-180). Amsterdam: John Benjamins.

Ranta, L. (2005). Language analytic ability and oral production in a second language: Is there a connection? In A. Housen & M. Pierrard (Eds.), *Investigations in instructed second language acquisition* (pp. 99-130). Berlin: Mouton de Gruyter.

Robinson, P. (2005). Aptitude and second language acquisition. *Annual Review of Applied Linguistics, 25*, 46-73

Roehr-Brackin, K. (2015). Long-term development in an instructed adult L2 learner: Usage-based and complexity theory applied. In T. Cadierno & S. W. Eskildsen (Eds.), *Usage-based perspectives on second language learning* (pp. 181-206). Berlin: de Gruyter.

Roehr-Brackin, K. (2018). *Metalinguistic awareness and second language acquisition*. London: Routledge.

Sawyer, M., & Ranta, L. (2001). Aptitude, individual differences, and instructional design. In P. Robinson (Ed.), *Cognition and second language instruction* (pp. 319-353). Cambridge: Cambridge University Press.

Skehan, P. (1998). *A cognitive approach to language learning*. Oxford: Oxford University Press.

Skehan, P. (2002). Theorising and updating aptitude. In P. Robinson (Ed.), *Individual differences and instructed language learning* (pp. 69-94). Amsterdam: John Benjamins.

Sparks, R., & Ganschow, L. (2001). Aptitude for learning a foreign language. *Annual Review of Applied Linguistics, 21*, 90-111.

Suárez, M. d. M., & Muñoz, C. (2011). Aptitude, age and cognitive development: The MLAT-E in Spanish and Catalan. In L. Roberts, G. Pallotti, & C. Bettoni (Eds.), *EuroSLA Yearbook* (Vol. 11, pp. 5-29). Amsterdam: John Benjamins.

Sáfár, A., & Kormos, J. (2008). Revisiting problems with foreign language aptitude. *IRAL, 46*, 113-136.

Tellier, A. (2013). Developing a measure of metalinguistic awareness for children aged 8-11. In K. Roehr & G. A. Gánem-Gutiérrez (Eds.), *The metalinguistic dimension in instructed second language learning* (pp. 15-43). London: Bloomsbury.

Tellier, A., & Roehr-Brackin, K. (2013). The development of language learning aptitude and metalinguistic awareness in primary-school children: A classroom study. *Essex Research Reports in Linguistics*, 62(1), 1-28.

Tellier, A., & Roehr-Brackin, K. (2017). Raising children's metalinguistic awareness to enhance classroom second language learning. In M. d. P. García Mayo (Ed.), *Learning foreign languages in primary school: Research insights* (pp. 22-48). Bristol: Multilingual Matters.

Wen, Z., Biedroń, A., & Skehan, P. (2017). Foreign language aptitude theory: Yesterday, today and tomorrow. *Language Teaching, 50(1)*, 1-31.

Table 1. Descriptive statistics MLAT-E

	Range	Min.	Max.	Mean	SD	Max. possible	Mean % correct
MLAT-E (pre-test)	98	15	113	69.98	24.469	130	53.83
MLAT-E (post-test)	99	23	122	86.50	23.35	130	66.54

Table 2. Descriptive statistics MLA test

	Range	Min.	Max.	Mean	SD	Max. possible	Mean % correct
MLA (pre-test)	42	10	52	27.52	9.105	83	33.16
MLA (post-test)	51	11	62	34.54	11.622	83	41.46

Table 3. Descriptive statistics L2 French proficiency test

	Range	Min.	Max.	Mean	SD	Max. possible	Mean % Correct
Pre-test	42	7	49	25.05	8.854	82	30.55
Immediate post-test	62	12	74	42.92	13.318	82	52.34
Delayed post-test	61	13	74	42.77	12.566	82	52.16
Combined post-test	61.5	12.5	74	42.85	12.405	82	52.26
French gains	60.5	-15.5	45	17.79	9.83	n/a	n/a

Table 4. Correlations between MLAT-E subtest scores and L2 French subtest gains

		Grammar	Listening	Writing	Reading
MLAT-E total	rho	.472**	.357**	.162	.531**
	p (2-tailed)	.000	.000	.089	.000
Number Learning	rho	.331**	.218*	.124	.313**
	p (2-tailed)	.000	.022	.195	.001
Finding Rhymes	rho	.411**	.353**	.112	.475**
	p (2-tailed)	.000	.000	.240	.000
Hidden Words	rho	.316**	.273**	.124	.354**
	p (2-tailed)	.001	.004	.194	.000
Matching Words	rho	.408**	.324**	.131	.503**
	p (2-tailed)	.000	.017	.171	.000

Note: ** = significant at the .01 level.

Table 5. Regression model: Prediction of L2 French gains

Predictors	Standardised coefficients: Beta	t	Sig.	Correlations (Part)	R^2 (% unique variance explained)
Number Learning	.170	1.765	.080	.142	2
Finding Rhymes	.305	2.413	.018*	.194	4
Hidden Words	-.126	-.976	.331	-.079	.6
Matching Words	.316	3.080	.003**	.248	6

Note: * = significant at the .05 level; ** = significant at the .01 level.