

Language learning aptitude as a predictor of late-life L2 learning at beginner level

Language Teaching Research

1–30

© The Author(s) 2025



Article reuse guidelines:

sagepub.com/journals-permissions

DOI: 10.1177/13621688251352260

journals.sagepub.com/home/ltr

**Karen Roehr-Brackin** 

University of Essex, UK

Renato Pavlekovic 

University of Essex, UK

Abstract

Recent work in the field of second language (L2) learning and teaching has aimed for improved representativeness by including older adult participants. Findings to date suggest not only that it is perfectly possible to learn a new L2 late in life, but also that, compared with younger samples, third-age learners' success may be less dependent on the nature of the instructional approach they are exposed to. Whereas the predictive power of language learning aptitude in young adults' instructed L2 learning has been amply demonstrated, we know very little about language aptitude as a predictor of late-life learners' L2 achievement. The present study addressed these issues by comparing the effectiveness of an explicit and an incidental instructional condition at the earliest stage of L2 learning. Volunteers ($n=80$) aged 60–83 completed the LLAMA aptitude battery and a serial reaction time task and participated in a suite of online language lessons targeting adjective-noun gender agreement in beginner-level Croatian. Our results show that the LLAMA tests significantly predicted L2 attainment. Aptitude components played a greater role in the incidental than in the explicit condition, indicating that the latter was cognitively less demanding. Nevertheless, participants were equally successful in the two conditions. The incidental group responded faster to posttest items throughout, and participants performed better on written than auditory items regardless of instructional condition. Taken together, these findings suggest that input modality may be more relevant for older adults than instructional approach. Participants' occupational status (working vs. retired) and self-concepts, including their confidence in themselves and their knowledge, emerged as important factors, highlighting the link between (meta)cognitive and socioaffective variables in late-life learners.

Keywords

Explicit learning, incidental learning, language learning aptitude, late-life learners, older adults, second language learning

Corresponding author:

Karen Roehr-Brackin, Department of Language and Linguistics, University of Essex, Wivenhoe Park, Colchester, CO4 3SQ, UK.

Email: kroehr@essex.ac.uk

I Introduction

Research into additional or second language (L2) learning in later life is no longer a minority pursuit. Applied linguists have realized that in view of ageing populations worldwide, it is important to understand how older adults learn language(s). Several studies on the topic have shown conclusively that successful L2 learning by participants aged 60+ is entirely feasible (e.g., Kliesch et al., 2018; Mackey & Sachs, 2012). Although certainly worth demonstrating empirically, this finding is ultimately unsurprising: despite age-related changes, the human brain retains plasticity throughout the lifespan and thus its capacity for learning new skills (for a recent overview of the field, see Cox & Sanz, 2023). Accordingly, the research agenda has been refined in recent years, yielding two focus points: the role of individual learner differences and the relative effectiveness of different instructional approaches. The present study contributes to this agenda by providing insights into two as yet under-researched issues: (1) the predictive power of language learning aptitude in older adults' L2 learning and (2) the relative effectiveness of a more vs. a less explicit approach to L2 instruction and learning.

II Background

I Language learning in older adults: Theoretical positions

Interest in L2 learning in later life has at least partly been fuelled by the potential usefulness of language learning as an activity that may stave off age-related cognitive decline and hence contribute to the building of cognitive reserve (Kliesch et al., 2018; Ramírez Gómez, 2016; Singleton & Pfenninger, 2018). It is well established that as we age, some of our cognitive and perceptual capacities weaken, with a gradual and in practice often imperceptible decline extending over many years. Processing speed and declarative memory (Kliesch et al., 2018; Ramírez Gómez, 2016; Singleton, 2018), working memory including the executive functions of attention and inhibition (Kliesch et al., 2018; Ramírez Gómez, 2016; Singleton & Ryan, 2004), and auditory acuity (Kliesch et al., 2018; Singleton & Ryan, 2004) are all affected in healthy ageing. At the same time, mature individuals exhibit strengths that are attributed to life experience. In particular, older learners may show high levels of awareness (Gabrys-Barker, 2018) and strategic expertise (Oxford, 2018; Piechurska-Kuciel & Szyszka, 2018; Singleton, 2018), while verbal and general knowledge remain stable or even increase in older adulthood (Oxford, 2018; Singleton, 2018; Singleton & Ryan, 2004).

In line with the predominant argument in cognitive psychology (e.g., Antoniou et al., 2013; Lindenberger, 2014; Salthouse, 2004), L2 research has tended to espouse a deficit view of healthy ageing. More recently, however, a reversal of the cause-effect equation has been proposed, based on the argument that observed changes in information processing capacity may be due to older individuals' greater experience and the associated increase of demands on memory retrieval (Pfenninger & Polz, 2018). In other words, it is growth in accumulated knowledge that results in a reduction of power and/or speed, not a decline in power and/or speed that triggers compensatory performance elsewhere. With this backdrop, researchers have called not only for a more positively framed

discourse surrounding healthy ageing (Roehr-Brackin, 2023), but also for furthering our understanding of which approaches to L2 learning and teaching may be particularly beneficial for older adults (Cox & Sanz, 2023).

2 Effectiveness of different instructional approaches for late-life L2 learning

In view of the increasing number of research studies into additional language learning by so-called third-agers, generally healthy individuals from around age 60 onwards (Kliesch & Pfenninger, 2021; Oxford, 2018), it is somewhat surprising that there are virtually no investigations into the effectiveness of different instructional approaches. Indeed, work on the impact of different teaching approaches has focused on the populations that are already over-represented in the field, that is, young adults in tertiary and adolescents in secondary education (Goo et al., 2015; Norris & Ortega, 2001), and to a lesser extent children in primary education (Roehr-Brackin, 2024).

At a theoretical level, a distinction can be made between explicit learning, knowledge, and teaching on the one hand and implicit learning, knowledge, and teaching on the other hand. Explicit learning is “characterized by the learner’s conscious and deliberate attempt to master some material or solve a problem” (Dörnyei, 2009, p. 136). The learner may try to identify systematic patterns, rules, or concepts that capture regularities in the input; for this endeavour, effort and strategic expertise are required. By contrast, implicit learning is an unconscious process of induction resulting in intuitive knowledge (N.C. Ellis, 1994). There is no conscious attempt to learn and no awareness of learning.

Explicit knowledge is knowledge that we are consciously aware of and that we can articulate in a verbal statement (R. Ellis, 2004). It is knowledge that can be called up on demand (Dörnyei, 2009) and includes semantic knowledge in the sense of the form-meaning pairings underlying vocabulary as well as knowledge of structural patterns typically described by pedagogical grammar rules. Conversely, implicit knowledge is intuitive knowledge that cannot be brought into awareness or articulated (Dörnyei, 2009; Hulstijn, 2005).

Teaching can be considered explicit “if rule explanation comprise[s] any part of the instruction [...] or if learners [are] directly asked to attend to particular forms and to try to arrive at metalinguistic generalizations on their own” (Norris & Ortega, 2001, p. 167). The latter type of explicit teaching is typically referred to as inductive, while the former type is described as deductive (Hulstijn, 2005). Conversely, instruction is implicit “when neither rule explanation nor directions to attend to particular forms [are] part of the treatment” (Norris & Ortega, 2001, p. 167).

While teachers and/or researchers can determine the instructional approach, they cannot control the learning processes that result from it. Thus, although explicit teaching is likely to encourage explicit learning and result in explicit knowledge, this is not guaranteed. By the same token, implicit instruction may encourage implicit learning, but (some) learners may nonetheless employ explicit (inductive) processes with a view to deriving explicit knowledge, for instance.

Whereas the relative effectiveness of more or less explicit instruction has been studied widely in younger populations (Goo et al., 2015; Norris & Ortega, 2001), the issue is

only beginning to be addressed in work with third-age learners. A series of laboratory-based studies into the acquisition of a morphosyntactic structure of Latin by older adults in the United States examined the impact of explicit information about the targeted L2 feature (Cox, 2017; Cox & Sanz, 2015) and the relative explicitness of corrective feedback (Lenet et al., 2011). Collectively, the research confirmed once more that older adults can successfully acquire new morphosyntactic structures. Comparing the performance of older with younger adults, less explicit accuracy-focused feedback was found to be more effective for older learners (ages 66–81), whereas more explicit feedback comprising both information on accuracy and a metalinguistic explanation worked better for younger learners (ages 18–21) (Lenet et al., 2011).

Focusing on a bilingual subsample, the researchers report that while the older adults achieved high mean scores on an explicit grammar test, they were significantly outperformed by the young adults who scored at ceiling (Cox & Sanz, 2015). When comparing bilingual with monolingual older adults, it was found that providing explicit information about the target structure prior to input-based practice yielded an advantage for bilingual but not monolingual participants. However, it had little effect overall (Cox, 2017): a finding that contrasts with the claim that explicit learning mechanisms become more prevalent in adults with advancing age (Ramírez Gómez, 2016) and the conjecture that, anecdotally at least, older learners desire grammar rules, metalinguistic explanations, and systematic analysis (Singleton & Ryan, 2004).

A recent study put these hypotheses to the test in an online classroom-based experiment to establish to what extent older L2 learners would show proficiency gains in an explicit vs. an implicit instructional condition (van der Ploeg et al., 2023). A total of 16 retired Dutch-speaking adults (mean age = 71.9) took part in a 3-month L2 English course delivered via Zoom. Learners were randomly assigned to either an implicit, meaning-focused or an explicit, form-focused instructional condition and followed a syllabus developed in accordance with a prior needs analysis. English proficiency was assessed at pretest, immediate posttest, and delayed posttest after a 3-month interval. Proficiency was operationalized via a monologic speaking task and verbal fluency, receptive vocabulary, and listening tests.

The researchers report that participants in both experimental conditions improved in equal measure in terms of oral proficiency on measures of grammar, lexis, and pronunciation. Vocabulary knowledge and verbal fluency likewise increased, especially at delayed posttest, while listening skills did not show any significant improvements. The only difference between the explicitly and implicitly instructed groups was found in oral accuracy, where the implicitly taught group made more mistakes in the beginning but showed a significant decrease in structural errors over time. The researchers conclude that, unlike younger learners, the older participants did not derive any advantages from explicit instruction. At the same time, the small sample size as well as evidence for substantial interindividual variation are acknowledged as potential limitations of the study (van der Ploeg et al., 2023).

The only other study we are aware of that examined different instructional approaches with older adult L2 learners outside a laboratory setting compared a monolingual with a multilingual teaching approach (Donnerer & Roehr-Brackin, in press). Participants aged 60–81 ($n = 50$) with either L1 German or L1 English took part in a 10-week course in

beginners' Italian, delivered online via Zoom. The monolingually taught group was exposed to L2 materials throughout, while approximately 20% of the learning materials in the multilingually taught group focused on different European languages to encourage metalinguistic reflection and cross-linguistic comparisons. The participants were pre-tested and posttested for L2 Italian knowledge and metalinguistic awareness. In order to account for individual learner differences, the LLAMA aptitude test suite was administered and information on background variables including prior language learning experience was gathered.

Participants in the two instructional conditions made significant progress in Italian, as expected, but their gains did not differ statistically. Thus, although the multilingual instructional approach did not convey any measurable advantages, it did not result in any disadvantages either, despite offering less L2 input and practice. Learners' performance at pretest was the strongest predictor of L2 achievement at posttest. Phonetic coding ability predicted L2 achievement and the acquisition of definite articles in particular: a target feature that was comparable to learners' L1s. Metalinguistic awareness and language-analytic ability predicted acquisition of *pro* drop: a target feature that has no equivalent in the learners' L1s and was thus deemed more challenging.

The researchers conclude that metalinguistic ability and the use of multilingual activities appear to be facilitative in older adults' L2 learning at beginner level, especially for experienced learners who have previously learned other (related) L2s, with metalinguistic awareness and L2 achievement mutually reinforcing each other (Donnerer & Roehr-Brackin, in press). Moreover, the aptitude components of phonetic coding ability and language-analytic ability made significant contributions to explaining participants' L2 achievement: a finding that aligns with accumulating evidence about the importance of individual difference variables as predictors of success in late-life language learning.

3 Individual differences in older adults' L2 learning

It has been argued that differences between individuals may become more marked with age because of longer and therefore potentially increasingly diverse life experiences (Gabrys-Barker, 2018; Kliesch & Pfenninger, 2021). As in younger learners, both cognitive and socioaffective factors are expected to interact with older learners' L2 development (Kliesch et al., 2018). Socioaffective factors include self-perceptions and self-esteem (Oxford, 2018), an individual's sense of purpose, and their overall wellbeing and motivation (Pfenninger & Polz, 2018). Cognitive factors have mostly been operationalized via measures of working memory and/or executive functions (Kliesch et al., 2018; Lenet et al., 2011; Mackey & Sachs, 2012).

In a groundbreaking longitudinal study, Kliesch and Pfenninger (2021) investigated the impact of a range of cognitive and socioaffective variables on L1 Swiss German older adults' achievement in L2 Spanish ($n=28$, ages 64–74). Participants completed L1 verbal fluency, simple and complex working memory, alertness, divided attention, and inhibition tasks as measures of cognitive capacity and single questions on overall wellbeing and L2 learning motivation as measures of socioaffect on a weekly basis over the course of 32 weeks. Instruction in Spanish comprised weekly classroom-based lessons supplemented by independent app-based practice on Duolingo. Development in Spanish

was assessed by means of a C-test, receptive tests of lexicon and grammar, and short oral interviews coded for fluency, targetlike use of vocabulary, lexical richness, and morpho-syntactic accuracy.

Unsurprisingly, the study yielded a rich and complex pattern of results. For the present line of argument, suffice it to say that the role of individual differences was shown to be of utmost importance: more often than not, mean group development was not representative of individual developmental trajectories. Moreover, significant L2 improvement could be observed in the early stages of training, throughout the training, or not at all, but delayed gains in later stages of training were extremely rare (Kliesch & Pfenninger, 2021).

Aside from Donnerer and Roehr-Brackin (in press) as summarized above, no published study on late-life L2 learning has included measures of language learning aptitude to assess cognitive individual differences, even though aptitude has been shown to be a crucial predictor of achievement in younger adults (Li, 2015, 2016; Wen et al., 2023). This gap in the literature has been pointed out elsewhere (Ramírez Gómez, 2016), and a recent investigation (Roehr-Brackin et al., 2023) began to address it by examining the suitability of existing aptitude measures for older adults and the relationship between aptitude for explicit and aptitude for implicit learning in such a population. The study thus took into account current conceptualizations of the construct according to which we can distinguish between capacities for explicit and implicit learning (Granena, 2020; Li, 2022; Pavlekovic & Roehr-Brackin, 2024).

The study also explored the role of demographic and socioaffective variables, examining to what extent participants' occupational status, chronological age, level of multilingualism, self-concept, emotional state, and leisure activities were associated with performance on the aptitude measures. A total of 64 volunteers aged 61–79 completed the LLAMA aptitude suite (Meara, 2005; Meara & Rogers, 2019) and a probabilistic serial reaction time (SRT) task (Kaufman et al., 2010). Results revealed that both the LLAMA and the SRT task proved challenging for the participants. The hypothesized distinction between implicit and explicit aptitude was supported empirically in that associative memory (LLAMA B) and language-analytic ability (LLAMA F) loaded on an explicit aptitude factor, whereas auditory pattern recognition (LLAMA D) and implicit sequence learning ability (SRT) loaded on an implicit aptitude factor, reflecting results obtained in a previous study with younger participants (Granena, 2013).

Interestingly, it was found that retired participants were at a disadvantage on implicit aptitude measures compared with participants who were still working; importantly, age was controlled for in this analysis. Furthermore, level of multilingualism in the sense of both quantity (number of languages) and quality (proficiency) of prior language learning experience and a more positive self-concept in terms of memory and cognition were associated with better performance on the implicit aptitude measures. Older participants within the age range of the sample were disadvantaged on explicit but not on implicit aptitude measures: a finding that appears to contradict the claim that implicit probabilistic learning abilities decline as we age in the same way as explicit processes do (Cox & Sanz, 2023). Instead, SRT task performance was found to be positively correlated with level of multilingualism, indicating better implicit sequence learning ability by participants with more extensive prior language learning experience.

The researchers conclude that the aptitude measures that were employed in the study were suitable for use with older adults, with the caveat that the speed of stimuli presentation on the SRT task should be reduced slightly. As the next step, an investigation into the predictive power of the aptitude tests in the context of an experimental research design was recommended (Roehr-Brackin et al., 2023). This is the focus of the present study.

III Research issues and questions

As outlined in the preceding sections, research into late-life language learning has become a vibrant subfield of applied linguistics. While existing studies have shown conclusively that L2 learning is possible at any age, only two studies have examined the relative effectiveness of different instructional approaches in third-age learners (Donnerer & Roehr-Brackin, *in press*; van der Ploeg et al., 2023), so further investigation is urgently needed (Cox & Sanz, 2023). Work to date on the effectiveness or otherwise of more or less explicit teaching and learning conditions has yielded contrasting hypotheses (Ramírez Gómez, 2016; Singleton & Ryan, 2004) and results (Cox, 2017; Lenet et al., 2011; van der Ploeg et al., 2023), pushing this issue to the top of the agenda.

Investigations into the role of intraindividual (Pfenninger & Kliesch, 2023) and inter-individual differences in late-life language learning (Kliesch et al., 2018; Kliesch & Pfenninger, 2021) have confirmed the relevance of both cognitive and socioaffective factors. Studies to date have operationalized cognitive abilities in terms of working memory and/or executive functions (Kliesch & Pfenninger, 2021; Kliesch et al., 2018), while the impact of language learning aptitude has been neglected. Measures of aptitude have featured in only two studies, one correlational (Roehr-Brackin et al., 2023), the other experimental (Donnerer & Roehr-Brackin, *in press*). However, no study to date has examined the predictive power of aptitude for both explicit and implicit learning in older adults, in accordance with current theorizing of the construct.

Accordingly, the present study addressed the following research questions:

- (1a) How do older adults aged 60+ perform on measures of aptitude for explicit and implicit learning?
- (1b) What is the relationship between the hypothesized components of aptitude for explicit and implicit learning?
- (2) How do individual differences in age, occupational status, level of multilingualism, and self-concept relate to performance on the aptitude measures?
- (3a) To what extent do aptitude for explicit and implicit learning predict older adults' achievement on a short online language course at beginner level in (a) an explicit learning condition and (b) an incidental learning condition?
- (3b) Do the two learning conditions lead to equivalent outcomes?

IV Methodology

We addressed these research questions in a quasiexperimental study during which volunteers learned the beginnings of a new language in the context of four online lessons in one of two learning conditions: explicit or incidental. Learning success was assessed by

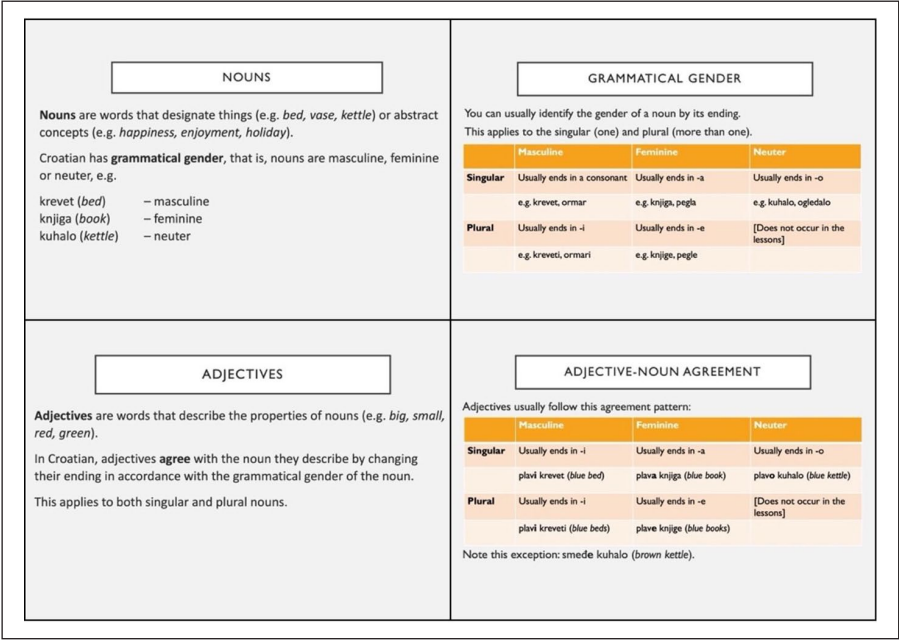


Figure 1. Adjective-noun gender agreement in Croatian.

means of an immediate posttest. In addition, we measured participants’ aptitude for explicit and implicit learning and gathered background information.

1 Learning conditions and materials

We developed a suite of lessons in beginners’ Croatian with adjective-noun gender agreement as the target feature. The rationale for our choice of Croatian was twofold. As languages other than English are under-represented in the field, we sought to redress the balance. Moreover, and connected with the previous point, we chose a language that is not widely learned as an L2 outside the Balkan region, so it would be new to our participants. Both progress in the L2 and effects of individual learner differences are most readily observable at early stages (Kliesch & Pfenninger, 2021; Ware et al., 2017). Opting for a completely unfamiliar L2 allowed us to keep the instructional treatment relatively short while still ensuring its effectiveness in the sense of measurable learning outcomes.

Croatian has three grammatical genders, with adjectives inflecting in accordance with the gender (and case) of the noun they qualify. Adjective-noun gender agreement is highly frequent and could therefore be incorporated into beginner-level lessons. The full inflectional paradigm is shown in Figure 1.

The instructional treatment comprised four 30-minute language lessons based around a storyline of a couple unpacking their belongings after moving house (Lesson 1a), shopping for groceries (Lesson 1b), getting further belongings delivered (Lesson 2a), and

purchasing household items and additional furniture (Lesson 2b). We aimed to make the lessons as entertaining and as ecologically valid as possible to approximate real-life language lessons, especially in the context of app-based language learning.

Lessons 1a and 2b comprised auditory input and practice, while Lessons 1b and 2a comprised written input and practice. The learning materials were developed as a series of PowerPoint slides featuring presentation and controlled practice activities in the L2, preceded by introductory slides that set the scene and drove the plot. The introductory slides were bilingual and bimodal, with Croatian speech and writing followed by an English translation in both speech and writing. All materials were receptive, i.e., participants listened to or read the L2 (presentation) and responded to four-way multiple-choice items (controlled practice); accuracy feedback was provided. Figure 2 shows example slides from Lessons 1b and 2b (The full set of presentation slides is available at https://osf.io/ub3a2/?view_only=0d853ae86a494b9da5b55f1d34fb9fbc).

The explicit learning condition included metalinguistic information in English about the target feature, that is, Lessons 1a and 2a began with the four slides shown in Figure 1. The incidental learning condition did not include any metalinguistic information or prompts to focus on form.¹ Instead, there were 50% more practice slides per lesson to ensure that participants encountered enough tokens to allow them to learn the inflectional paradigm. Otherwise, the materials were identical. Within the time limit of each lesson, participants could move freely through the material.

Prior to engaging in Lessons 1a and 2a, participants went through 25-minute vocabulary learning sessions in order to familiarize themselves with the lexical items featuring in the subsequent lessons. Vocabulary was presented both auditorily and in writing with English translation equivalents and, whenever possible, pictorial support to facilitate learning and subsequent recall. Within the given time limit, participants could move freely through the material.

Following each vocabulary learning phase, participants completed a vocabulary test that they had to pass at a minimum of 80% accuracy in order to proceed to the lessons. If a participant did not reach this criterion, they were allowed to return to the learning phase and then attempt the test again; up to three attempts were permitted. No participants were excluded from the study due to failing a vocabulary test three times.

After completion of the language lessons, participants took a posttest assessing their learning of the target feature. The test comprised 37 four-way multiple-choice items whose format and modality matched those of the practice items that had been encountered during the lessons (see Figure 2). No feedback was provided. Reliability was good: Cronbach's $\alpha = .80$ (full set of practice and posttest items available at https://osf.io/qp69v/?view_only=e11fe063d57343bbaf1d25052ae3f4ce).

2 Cognitive ability measures: aptitude for explicit and implicit learning

Language learning aptitude was measured by means of the LLAMA test suite (<https://llamatests.org/>). We used v.3 (Meara & Rogers, 2019; Rogers et al., 2023), which comprises four subtests: LLAMA B for associative memory of novel word-referent combinations, LLAMA D for auditory pattern recognition based on spoken words in an unknown

W
R
I
T
T
E
N

PRESENTATION SLIDES



PRACTICE SLIDES

Ovo je  slika

☐ plava

☐ plavo

☐ plave

☐ plavi

Ovo je 

☐ narančasta ribež

☐ roza sat

☐ rozi sat

☐ narančasti ribež

A
U
D
I
T
O
R
Y

PRESENTATION SLIDES



PRACTICE SLIDES

Question 

☐ a

☐ b

☐ c

☐ d

Question 

☐ a

☐ b

☐ c

☐ d

Figure 2. Example slides from Lessons 1b (written) and 2b (auditory).

language, LLAMA E for sound-symbol association, and LLAMA F for grammatical inferencing/language-analytic ability based on an artificial minilanguage (see Figure 3).

LLAMA B required participants to learn 20 word-referent combinations within 2 minutes followed by an untimed test phase. The maximum score was 20.

LLAMA E required the learning of 24 one-syllable sound-symbol combinations during a 2-minute learning phase, followed by an untimed test phase where participants selected from a display 20 two-syllable combinations they heard spoken. We applied a partial-credit scoring scheme because it led to higher reliability than dichotomous scoring: participants were awarded a point for each correctly recalled syllable in its correct position, resulting in a maximum score of 40.

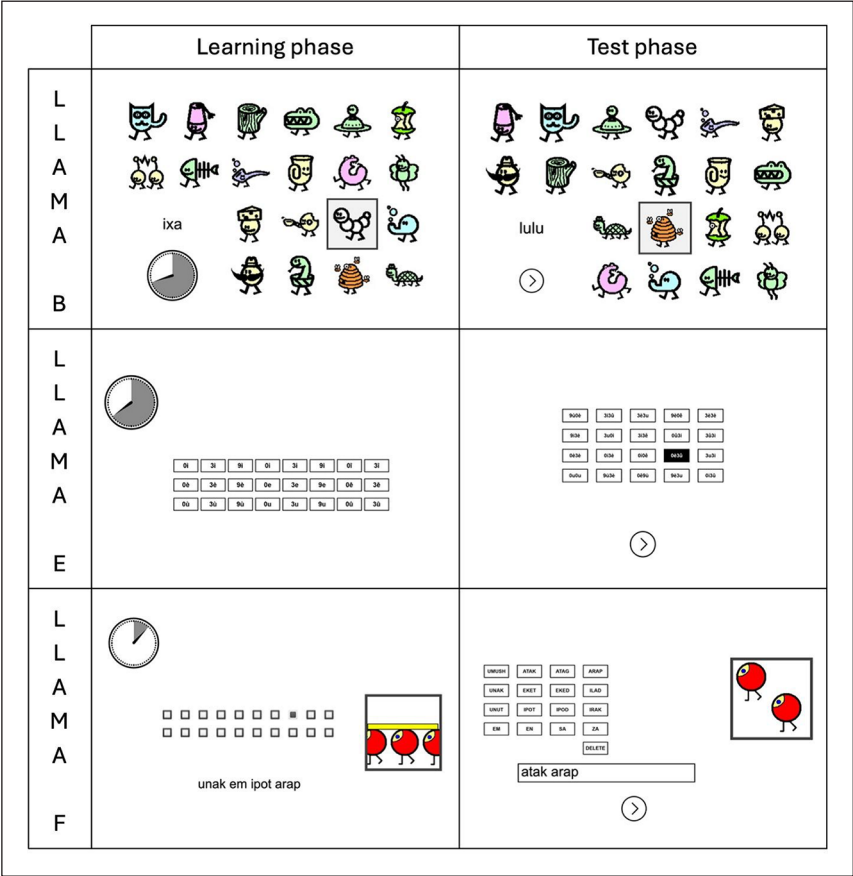


Figure 3. Interfaces of LLAMA B, E, and F.

LLAMA F required participants to work out the regularities underlying an artificial minilanguage within 4 minutes. The untimed test phase assessed responses based on rules of word order, singular/plural markers, shape and colour of depicted objects, prepositions, and adjectival agreement. Again, we applied a partial-credit scoring scheme because of increased reliability compared with dichotomous scoring. Points were awarded for each target feature that had been worked out correctly, resulting in a maximum score of 120. LLAMA B, E, and F are measures of aptitude for explicit learning.

LLAMA D in v.3 has a test phase only. Participants listen to a word spoken in an unknown language and indicate via a yes/no button whether or not they have heard the word before. There were 50 items in total: 10 target items that occurred three times each, interspersed with 20 distractor items. The maximum score was 50. Given the absence of a learning phase, LLAMA D has been hypothesized to measure aptitude for implicit learning, although empirical results to date are ambiguous in this regard (Granena, 2013; Li & DeKeyser, 2021; Pavlekovic & Roehr-Brackin, 2024; Roehr-Brackin et al., 2023).

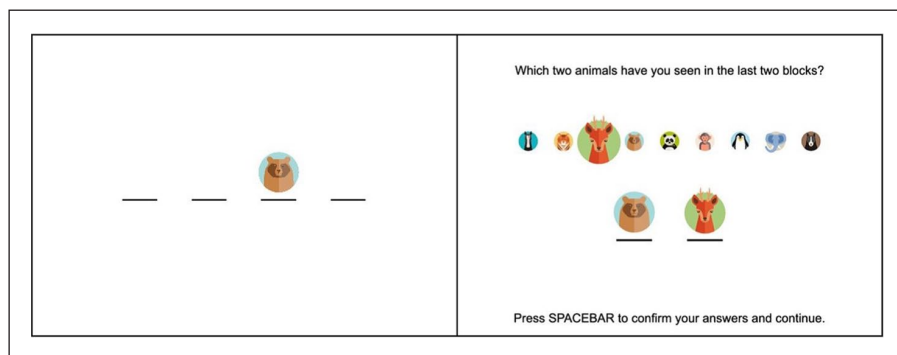


Figure 4. Serial reaction time task interface.

We used a probabilistic SRT task (Kaufman et al., 2010) to measure implicit sequence-learning ability as a component of aptitude for implicit learning. Participants responded to visual stimuli presented in one of four screen locations. The sequence in which the stimuli appeared alternated pseudorandomly between two patterns, a training sequence (85% of the time) and a control sequence (15% of the time). There were eight blocks of 120 trials each, preceded by 30 practice trials.

Following the recommendations in Roehr-Brackin et al. (2023), we made two adjustments to the task. First, we increased the interval at which stimuli were presented from 1000 to 1100 ms. Second, we changed the look of the stimuli by replacing black squares with the faces of cartoon animals to increase visual appeal and make the task less tedious for participants. In order to aid concentration, we asked participants every two blocks to indicate which two animals they had seen (see Figure 4). The changes led to the desired result, with almost all participants completing the task successfully. This compares favourably with just two-thirds of the sample producing usable data in Roehr-Brackin et al. (2023).

Inaccurate responses (pressing the wrong button, being timed out) accounted for 13% of the data. Reaction times that were 2.5 *SD* above each participant's mean within a given block were discarded, resulting in the removal of a further 2% of the data. A participant's SRT task score was determined by calculating the difference between reaction times on control and training trials in blocks 3 to 8, where learning effects were observable at group level.

3 Socioaffective and demographic measures

Demographic information including prior language learning experience, participants' self-concepts, and perceptions of the instructional materials were collected via an exit questionnaire (available at https://osf.io/epnf3/?view_only=e57ca7a532104954b76b32e77062673c). Participants indicated their perceptions of the learning materials by responding on a five-point agreement scale to 13 statements about time available for learning, entertainment value, difficulty of the materials, and preferred input modality. Reliability

Table 1. Study procedure.

	Length (minutes)	Task
Day 1	25	Vocabulary learning 1
	5	Vocabulary test 1
	30	Lesson 1a
	30	Lesson 1b
Day 2	25	Vocabulary learning 2
	5	Vocabulary test 2
	30	Lesson 2a
	30	Lesson 2b
	10	Posttest
Day 3	30	LLAMA
	30	Serial reaction time task
	15	Exit questionnaire

was very good: Cronbach’s alpha = .87. Self-concept was assessed by asking participants to report on their physical and mental state, abilities, and skills on a five-point agreement scale to 10 statements such as “I am fit,” “I have a good memory,” or “I am good at learning languages.” The items were the same as in Roehr-Brackin et al. (2023). Reliability was good: Cronbach’s alpha = .78.

4 Procedure

The study was conducted online, enabling participants to complete all components at times that were convenient to them. Previous research with late-life language learners has confirmed the feasibility of online data collection with this demographic (Piechurska-Kuciel & Szyszka, 2018; Roehr-Brackin et al., 2023; Ware et al., 2017). The learning materials could be accessed via our institution’s Moodle site, tests had been programmed into Psychopy and were administered via Pavlovia, and the exit questionnaire was provided on Qualtrics. Participants were automatically moved between these platforms as they completed the various stages of the study, as summarized in Table 1.

As indicated by Table 1, the study ran over 3 separate days that could be, but did not have to be, consecutive, though participants were asked to complete all components within a 7-day period. As reported above, the vocabulary learning sessions and lessons had fixed time limits. The questionnaire and tests were not timed, so the times listed in Table 1 are approximate.

Participants who completed all study components were offered a gift voucher worth GBP 20 as a token reward for their efforts. The study was approved by our institution’s Ethics Subcommittee 3 before participant recruitment began (reference ETH2324-0386).

5 Participants

We recruited participants who were aged 60 or over and proficient in English with no knowledge of Croatian or other Slavic languages. Furthermore, access to a laptop or

desktop computer with speakers or headphones and a stable internet connection was required.

A total of 80 volunteers completed all components of the study. They were aged between 60 and 83 (mean=70, median=69), with 58 identifying as female, 21 as male, and 1 preferring not to disclose their gender. Participants used to work ($n=61$) or were still working either full-time or part-time ($n=19$) in a range of professional fields: education/research (28), administration (17), accountancy/finance/IT/law (12), consultancy/management/translation/writing (12), health/social work (4), engineering (3), civil service (2), and the service industry/transport (2). The professional fields suggest a relatively privileged sample in terms of socioeconomic status, which is borne out by an overall high level of education. While reports of educational level ranged from no formal qualifications to a doctoral degree, there is a pronounced negative skew in the distribution, with a university degree the norm.

The majority of participants were L1 speakers of English ($n=67$). In addition, there were L1 speakers of French (3), German (3), Basque, Cantonese, Gujarati, Japanese, Mandarin, Spanish, and Swedish (1 each). Participants were asked to report all additional languages they had ever learned or tried to learn and to self-assess their level of proficiency in each language on a simple four-point scale. This resulted in between 1 and 6 additional languages beyond the L1 being reported (mean=3.8); overall, 35 different languages were listed. English was the most frequent L2 for non-L1 speakers of English. The other most frequently mentioned additional languages were French, German, Spanish, Italian, and Latin, with the latter again pointing towards high levels of education. Following the approach taken in Roehr-Brackin et al. (2023), we calculated a multilingualism score by adding up the z -scores for number of languages learned and mean proficiency level of languages. This score constitutes a combined measure of quantity (languages learned/encountered) and quality (self-assessed proficiency) of prior language learning and was used in the analyses reported below.

Participants had positive self-concepts, as shown in Figure 5. Collectively, the participants also reported very positive perceptions of the instructional materials, with only one outlier on the negative side of the scale (see Figure 5).

As volunteers came forward, they were allocated on an alternating basis to one of the two learning conditions. Since we could not predict who would complete the study and who would either withdraw or not start at all, we ended up with slightly uneven group sizes: $n=38$ in the explicit and $n=42$ in the incidental learning condition.

A comparison between the two groups on the individual difference variables measured in the present study yielded no significant differences in terms of age, level of education, level of multilingualism, or performance on LLAMA B, D, F, or the SRT task (all p -values $> .122$). However, the groups differed in terms of LLAMA E scores (Mann-Whitney U , $p=.013$) and self-concept (Mann-Whitney U , $p=.033$), with learners in the incidental condition outperforming learners in the explicit condition on LLAMA E and reporting more positive self-concepts (descriptive statistics available at https://osf.io/fjgdt/?view_only=9032486f4b124910882278238cb9f3a1). In terms of occupational status, the distribution was exactly the same, with each group comprising 76% retired and 24% working participants.

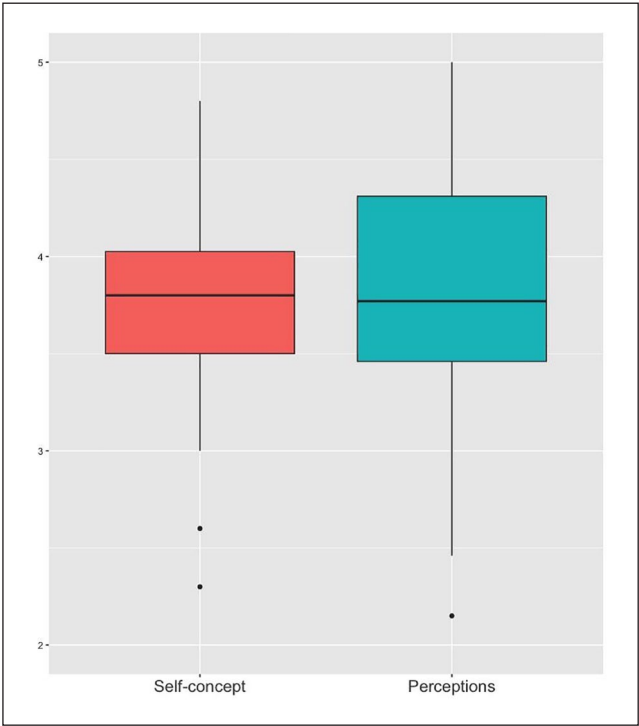


Figure 5. Self-concepts and perceptions of the learning materials.

V Results

In what follows, we address the research questions in order and outline the approach to data analysis as we proceed.

The first research question asked how older adults aged 60+ performed on the measures of language learning aptitude used in the study and what the relationship between the hypothesized components of aptitude for explicit and implicit learning was. Table 2 presents the descriptive statistics for the LLAMA test suite and the SRT task.

The LLAMA subtests exhibit good or excellent reliability, and standard deviations suggest suitable discrimination between test takers. The means and positive skew show that the LLAMA subtests were challenging for the participants, especially LLAMA B (associative memory), though the result is entirely comparable with that obtained in a previous study (Donnerer & Roehr-Brackin, in press). LLAMA D (auditory pattern recognition) scores are likewise comparable at around 30% (Donnerer & Roehr-Brackin, in press; Roehr-Brackin et al., 2023).

Participants’ performance on the SRT task is remarkably similar to the results reported in Roehr-Brackin et al. (2023), which is the only other study to have used this measure with late-life learners. The mean response time difference of 12.8 ms between training and control trials is statistically significant with a medium effect size

Table 2. Descriptive statistics: aptitude measures.

Subtest	<i>n</i>	<i>k</i>	Min	Max	<i>M</i>		<i>SD</i>		Skew	<i>S</i> – <i>W</i> (<i>p</i>)	α/r_{SB}
					Raw	Cor.	Raw	Cor.			
LLAMA B	80	20	0	20	4.31	21.56	3.79	18.94	1.56	.001	.82
LLAMA D	80	50	0	34	15.93	31.85	9.43	18.86	0.02	.019	.74
LLAMA E	80	20	2	40	19.09	47.72	11.20	28.00	0.273	.001	.92
LLAMA F	80	20	18	114	60.30	50.24	24.57	20.47	0.234	.093	.93
SRT	73	720	–288 ¹	272 ¹	12.80 ¹		17.79 ¹		–0.684	.054	.43

Note. *n* = number of participants; *k* = number of items/trials; α/r_{SB} = reliability for the LLAMA based on Cronbach's alpha and for the serial reaction time (SRT) task based on split-half with Spearman-Brown correction.
¹Milliseconds.

(following Plonsky & Oswald, 2014): $t = -6.008$, $df = 73$, $p < .001$, Cohen's $d = -0.698$. Thus, an implicit sequence learning effect is in evidence.

Correlations between the various aptitude measures can be found in Figure 6. There is a positive association of moderate to medium strength between the subtests measuring aptitude for explicit learning (henceforth: explicit aptitude), that is, LLAMA B (associative memory), E (sound-symbol association), and F (language-analytic ability).

We conducted a factor analysis (principal components extraction with oblimin rotation) on the aptitude measures, ascertaining that assumptions were met ($KMO = .640$, Bartlett's test of sphericity $< .001$). Two factors with an eigenvalue > 1 were extracted, explaining 62% of the variance. The factor loadings as shown in Figure 7 are unequivocal and confirm the hypothesized components of explicit (LLAMA B, E, F) and implicit (LLAMA D, SRT task) aptitude.

The second research question asked how individual differences in age, occupational status, level of multilingualism, and self-concept were related to performance on the aptitude measures. Correlations (see Figure 6) show that age (year of birth) is moderately correlated with LLAMA D (auditory pattern recognition) and E (sound-symbol association) and approaching significance for LLAMA F (language-analytic ability), indicating an advantage for younger participants within the 13-year age range of the sample. Moreover, self-concept is positively related to LLAMA D (auditory pattern recognition); that is, participants with a more positive view of their own health, happiness, and abilities performed better.

Level of multilingualism is weakly to moderately correlated with LLAMA B (associative memory) and E (sound-symbol association), indicating a link with more extensive prior language learning experience. Similarly, level of education is weakly to moderately correlated with LLAMA B (associative memory), E (sound-symbol association), and F (language-analytic ability), showing an association between educational background and performance on the explicit aptitude subtests. As level of multilingualism and level of education also correlate with each other, partial correlations were run, factoring out first one and then the other variable. This results in nonsignificant relationships with the explicit LLAMA subtests throughout (p -values $> .126$), although the partial correlations between level of education and LLAMA E (sound-symbol association)

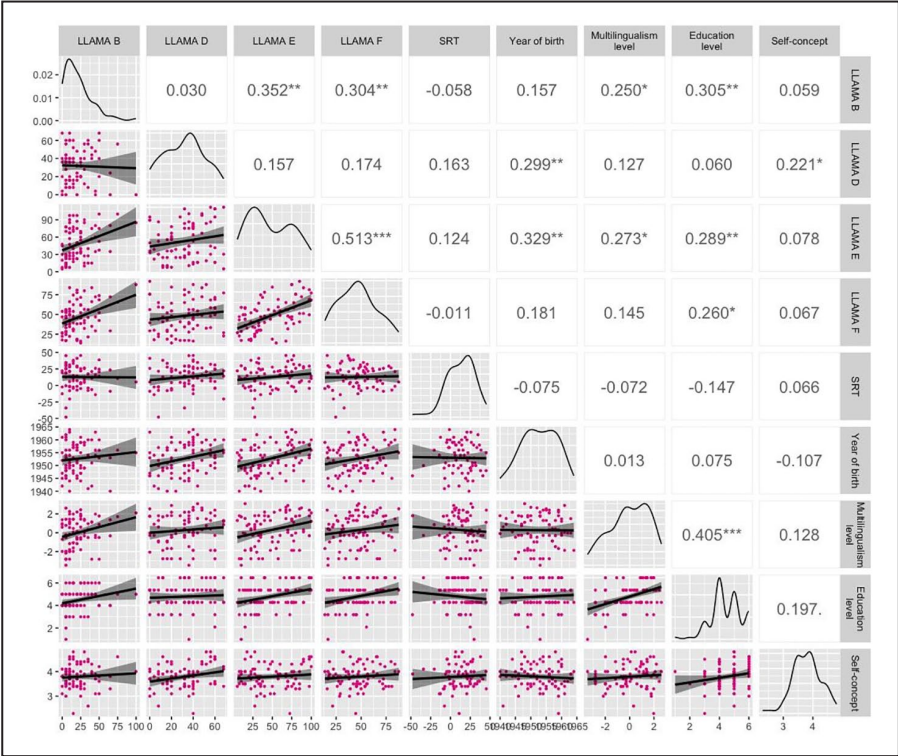


Figure 6. Correlations (Spearman's rho): aptitude measures, self-concept, and background variables.
SRT: serial reaction time task.

($p = .053$) and to a lesser extent LLAMA F (language-analytic ability) ($p = .090$) approach significance. Overall, this suggests that the two background variables are partially overlapping and mutually reinforce one another.

The final background variable to be scrutinized is occupational status. The sample included 61 retired individuals and 19 individuals who were still working, either full-time or part-time. In Figure 8, we can observe descriptive differences on the phonetic coding ability subtests (LLAMA D, auditory pattern recognition; and E, sound-symbol association), with the working participants seemingly at an advantage. On LLAMA F (language-analytic ability) and the SRT task, the retired group had a wider distribution of scores, while the working group was more tightly clustered. In terms of central tendency, the median scores were similar, however. Performance patterns on LLAMA B (associative memory) are similar as well.

To investigate the role of occupational status, we inferentially compared aptitude scores between retired and working individuals. Given the nonnormality of the data, we employed a permutational multivariate analysis of covariance (PERMANCOVA), with occupational status (retired vs. working) as the explanatory variable, aptitude scores as

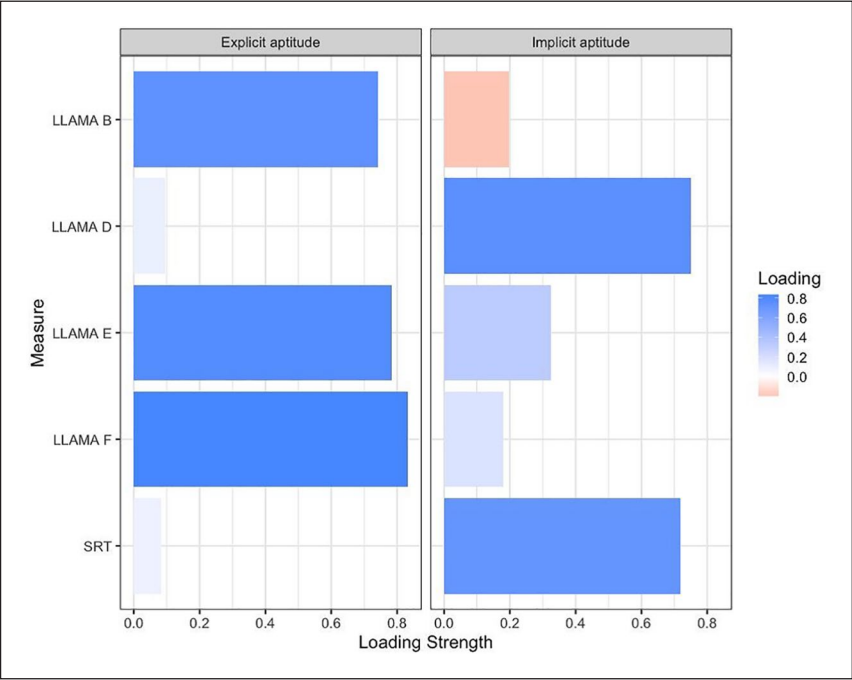


Figure 7. Factor loadings: aptitude measures.

the dependent variables, and age (year of birth) as a covariate, given that working participants were also younger. The model was fitted using the `adonis2()` function in the `vegan` package for *R* (version 2.6-8, <https://CRAN.R-project.org/package=vegan>). To mitigate the risk of multicollinearity, we used factor scores for explicit and implicit aptitude. The analysis was based on Euclidean distance, with 999 permutations.

We found a statistically significant effect of occupational status on combined aptitude scores ($F=3.12$, $R^2=.0385$, $p=.039$), accounting for just under 4% of the variance. Pairwise comparisons were run to assess the effect of occupational status on explicit and implicit aptitude separately while controlling for age. The results showed no significant effect of occupational status on explicit aptitude ($F=2.49$, $R^2=.0310$, $p=.109$) and a significant effect on implicit aptitude ($F=3.76$, $R^2=.0460$, $p=.04$). In other words, working participants outperformed retired individuals in terms of implicit aptitude when age was controlled for. This result replicates a previously reported finding (Roehr-Brackin et al., 2023). Figure 9 presents a side-by-side comparison of explicit and implicit aptitude scores for retired and working participants.

The third research question asked to what extent explicit and implicit aptitude would predict older adults’ achievement on a short online language course at beginner level in (a) an explicit learning condition and (b) an incidental learning condition, and whether the two learning conditions would lead to equivalent outcomes. Taking the latter question first, the descriptive statistics for posttest performance in the explicit condition are

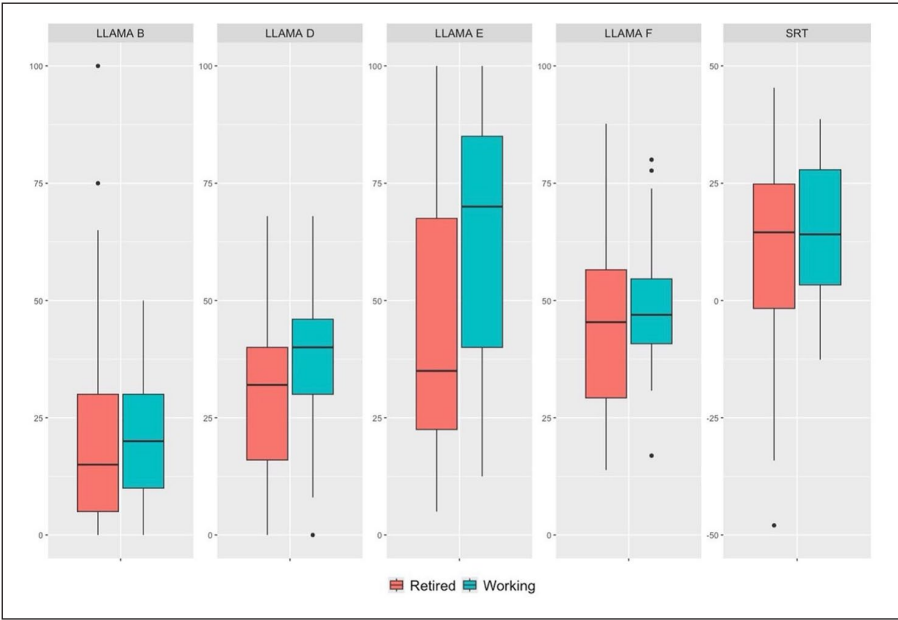


Figure 8. Performance of working and retired participants on the aptitude measures.

given in Table 3 and the descriptive statistics for posttest performance in the incidental condition are given in Table 4.

Overall posttest scores of above 80% accuracy are similar across the two instructional conditions, demonstrating that participants learned the target feature to a high level of success. Scores are likewise similar across learning conditions when written and auditory items are scrutinized separately. If we consider the average time it took participants to respond to the (untimed) posttest items, it is noteworthy that the incidental group was faster throughout. Inferentially, there are no statistical differences between the two groups in terms of accuracy, either for the posttest as a whole (Mann-Whitney U , $p = .406$) or written ($p = .262$) and auditory ($p = .516$) items separately. By contrast, the differences in response times are significant, with the incidental group responding faster across the board (overall $p = .009$, written items $p = .004$, auditory items $p = .017$).

Comparing across item types, participants achieved higher scores on written than on auditory items (Wilcoxon Signed Rank, $p < .001$). Moreover, responses to written items were faster than responses to auditory items ($p < .001$). Note that the latter result is likely to be a methodological artifact to at least some extent, given that all written multiple-choice options were visible at a glance, whereas participants had to separately click on each auditory multiple-choice option in order to listen, compare, and decide. Nevertheless, the difference in response times between written and auditory items is considerable and in line with the greater difficulty of the auditory modality, which may have prompted participants to listen to options more than once and thus spend (even) longer on auditory items.

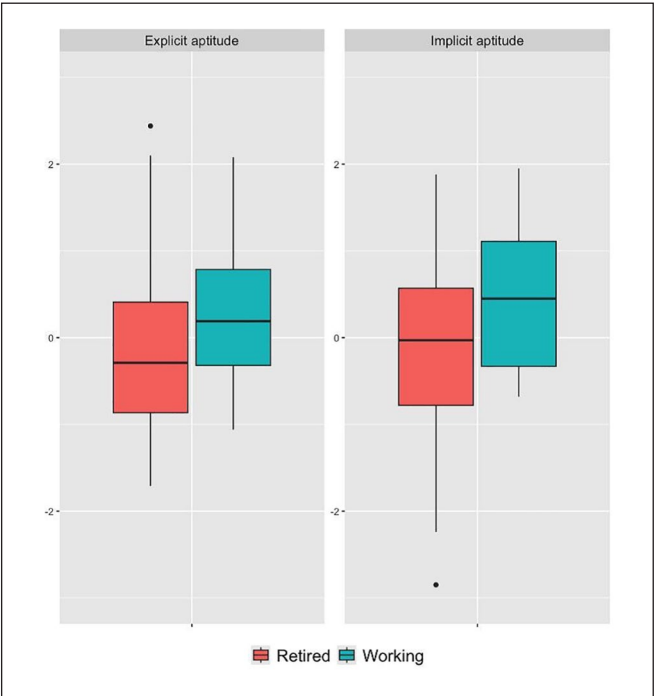


Figure 9. Performance of working and retired participants on explicit and implicit aptitude measures.

Table 3. Descriptive statistics: posttest performance in the explicit condition.

Measure	<i>n</i>	<i>k</i>	<i>M</i>		<i>SD</i>		Skew	<i>S</i> – <i>W</i> (<i>p</i>)
			Raw	Cor.	Raw	Cor.		
Posttest	38	37	31.03	83.85	4.40	11.88	–0.756	.001
Posttest RT	38	37	22.89		7.77		1.090	.009
Posttest W	38	16	14.79	92.43	1.45	9.09	–1.447	.001
Posttest W RT	38	16	17.65		7.26		1.294	.001
Posttest A	38	21	16.24	77.32	3.64	17.34	–0.824	.007
Posttest A RT	38	21	26.89		8.76		0.787	.062

Note. W = written condition; A = auditory condition; RT = reaction time; *n* = number of participants; *k* = number of items.

In order to establish the predictive power of the aptitude measures for posttest performance in the two instructional conditions, we first examined the correlational patterns in each group, as shown in Figures 10 and 11.

Table 4. Descriptive statistics: posttest performance in the incidental condition.

Measure	<i>n</i>	<i>k</i>	<i>M</i>		<i>SD</i>		Skew	<i>S</i> – <i>W</i> (<i>p</i>)
			Raw	Cor.	Raw	Cor.		
Posttest	42	37	30.14	81.47	4.88	13.18	–1.010	.003
Posttest RT	42	37	18.78		5.61		1.133	.004
Posttest W	42	16	14.26	89.14	2.01	12.58	–1.414	.001
Posttest W RT	42	16	13.75		4.74		1.864	.001
Posttest A	42	21	15.88	75.62	3.31	15.78	–0.705	.015
Posttest A RT	42	21	22.61		6.71		0.706	.084

Note. W = written condition; A = auditory condition; RT = reaction time; *n* = number of participants; *k* = number of items.

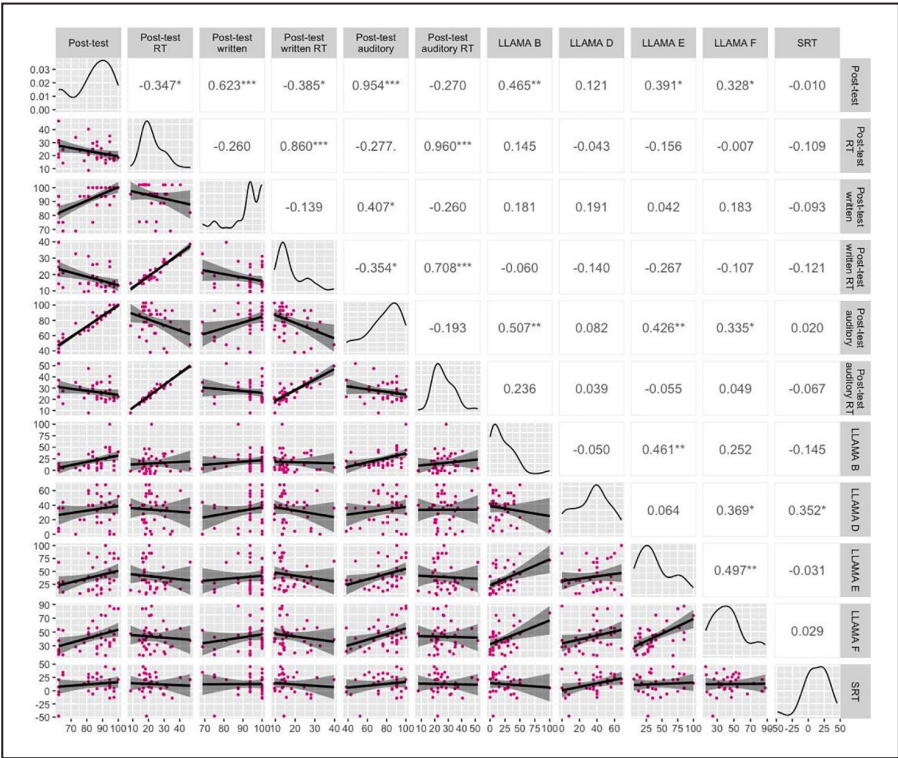


Figure 10. Correlations (Spearman's rho) between aptitude measures and posttest: explicit condition.

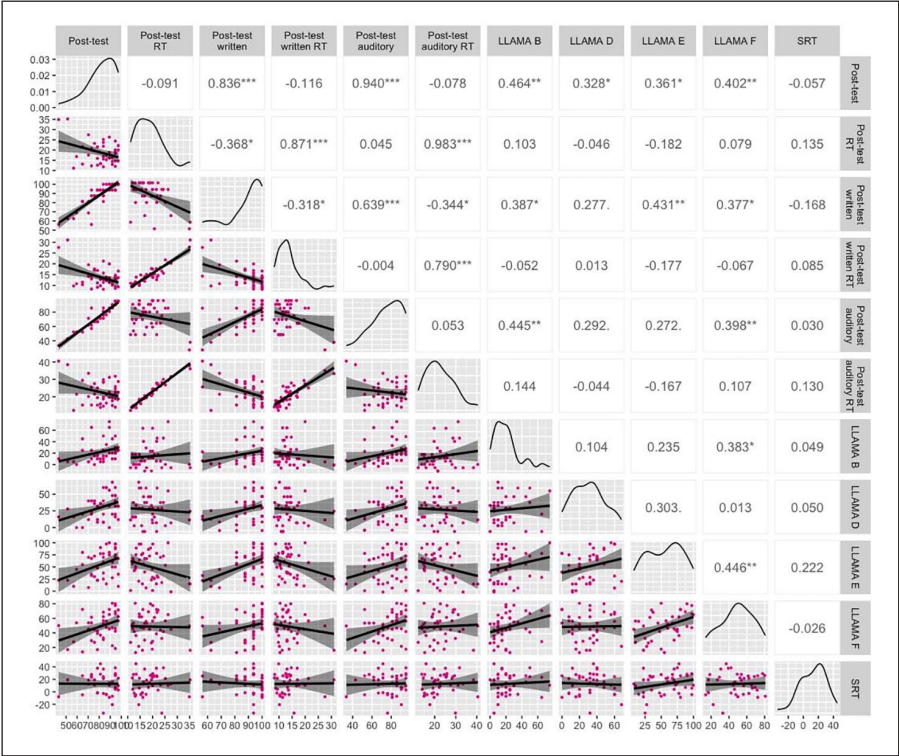


Figure 11. Correlations (Spearman's rho) between aptitude measures and posttest: incidental condition.

In the explicit group, a negative correlation between accuracy and response time demonstrates that faster responses were associated with better performance. While counter-intuitive at first glance, this result is less surprising if we bear in mind the length of the response times overall. As pointed out above, the posttest was not timed, so response times were leisurely across the board. In such a scenario, it is likely that a speedier response reflects greater confidence in one's answer, i.e., knowledge that could be put to use without much doubt or hesitation. Response times do not correlate with any of the aptitude measures, reinforcing the point that they may above all be indicative of confidence in acquired knowledge rather than of underlying ability. Overall posttest performance in terms of accuracy correlates with LLAMA B (associative memory) and E (sound-symbol association) scores, and this correlation is driven by the auditory items, that is, items in the more challenging modality.

The incidental group shows a different correlational pattern. For the posttest overall, there is no association between accuracy and response times, although we see a negative correlation between accuracy and response times on the written items. In other words, on written items only, a faster response coincided with better performance. As in the explicit group, there is no relationship between response times and any of the aptitude measures.

Table 5. Regression model: explicit condition.

Model	R ²	ΔR ²	ΔF	Sig. ΔF
LLAMA B	.172	.172	7.464	0.010
LLAMA B, LLAMA E	.198	.027	1.162	0.288
LLAMA B, LLAMA E, LLAMA F	.205	.006	0.269	0.608
LLAMA B, LLAMA E, LLAMA F, LLAMA D	.249	.045	1.971	0.170
LLAMA B, LLAMA E, LLAMA F, LLAMA D, SRT	.260	.010	.436	0.514

Table 6. Regression model: incidental condition.

Model	R ²	ΔR ²	ΔF	Sig. ΔF
LLAMA B	.100	.100	4.451	0.041
LLAMA B, LLAMA E	.164	.064	2.964	0.093
LLAMA B, LLAMA E, LLAMA F	.226	.062	3.063	0.088
LLAMA B, LLAMA E, LLAMA F, LLAMA D	.307	.081	4.333	0.044
LLAMA B, LLAMA E, LLAMA F, LLAMA D, SRT	.313	.006	0.310	0.581

Overall posttest performance in terms of accuracy is significantly associated with all LLAMA subtests, suggesting a greater role for aptitude in the incidental learning condition. The explicit subtests LLAMA B (associative memory), E (sound-symbol association), and F (language-analytic ability) correlate with accuracy on the written items, and LLAMA D (auditory pattern recognition) shows a trend in the same direction. LLAMA B (associative memory) and F (language-analytic ability) correlate with accuracy on the auditory items, with both LLAMA D (auditory pattern recognition) and E (sound-symbol association) showing trends in the same direction.

As a final step, we conducted a linear regression analysis (“Enter” method) in order to establish the predictive power of the aptitude measures for posttest performance in each of the learning conditions. The LLAMA subtests and the SRT task were entered as independent variables in order of correlational strength and regardless of significance. The resulting regression models are presented in Tables 5 and 6.

In the explicit group, LLAMA B (associative memory) significantly predicts posttest performance, accounting for 17% of the variance. In the incidental group, LLAMA B (associative memory) and D (auditory pattern recognition) are significant predictors, respectively accounting for 10% and 8% of the variance in posttest scores, while LLAMA F (language-analytic ability) and E (sound-symbol association) approach significance, each accounting for a further 6%. The overall trend thus suggests that the LLAMA suite collectively explains 30% of the variance in posttest scores in the incidental learning condition.

VI Discussion

The findings arising from the present study speak to two as yet unresolved issues in late-life language learning: first, the relative effectiveness of different instructional approaches

for third-age learners, and second, the role of aptitude in older adults' L2 learning. In the following, we discuss these issues.

1 Effectiveness of different learning conditions

In the present study, participants aged 60+ learned the beginnings of a new language via a suite of ecologically valid online language lessons. The participants experienced one of two instructional approaches: an explicit learning condition that included a full metalinguistic description of the targeted feature, and an incidental learning condition without metalinguistic information or prompts to focus on form but with additional practice opportunities. Our findings show that posttest performance in the explicit condition was predicted by LLAMA B (associative memory). This result suggests that the metalinguistic scaffolding that was available in this condition neutralized the need for high levels of language-analytic or phonetic coding abilities. The inflectional paradigm was presented in full, so success was above all dependent on memory, i.e., the ability to recall the various endings for the various adjectives and nouns. We can further conjecture that the significant role of explicit (declarative) memory is consistent with a lack of proceduralization of any learned knowledge and heavy reliance on explicit knowledge that is accessed via controlled processing, as one might expect after a relatively short period of exposure to L2 input.

Conversely, posttest performance in the incidental condition was associated with several aptitude components and predicted by all four LLAMA subtests (two of these marginally). Without metalinguistic scaffolding, not only associative memory but also language-analytic and phonetic coding abilities were called upon, since the target pattern had to be inferred from the input materials that were provided in both written and auditory format. Therefore, higher levels in all aptitude components were beneficial in this more demanding learning condition.

While it is perhaps unsurprising that the availability of metalinguistic information reduces the cognitive challenge when the learning target is an inflectional paradigm, posttest results did not show any statistical advantage for the explicit group. Instead, participants in the two learning conditions achieved similarly high scores, with a mean of over 80% accuracy. Two potentially complementary explanations suggest themselves. First, the instructional materials appear to have been pitched exactly right with the given balance of presentation and controlled practice in both written and auditory format, enabling our volunteers to successfully learn the target feature. The very positive perceptions reported by participants are further testimony to the quality of the materials.

Second, and arguably more importantly, our finding is in line with the results reported in previous research assessing the effectiveness of different instructional approaches outside a lab setting: in those studies too, no between-group differences in posttest performance were observed, regardless of whether an explicit and an implicit approach (van der Ploeg et al., 2023) or a monolingual and a multilingual approach (Donnerer & Roehr-Brackin, in press) were compared. While it is admittedly still early days, the emerging picture at this point suggests that older adults may be particularly autonomous learners who are able to find their own way and succeed in equal measure in different

instructional contexts (for a related argument, see Donnerer & Roehr-Brackin, under review). Having said this, it is worth bearing in mind that participants in the incidental condition had higher LLAMA E (sound-symbol association) scores and reported more positive self-concepts, both of which may have contributed to better performance.

The set-up of the present study allowed us to examine two further issues that provide interesting insights into older adults' L2 learning: input modality and response time. Our participants experienced both auditory and written input and, accordingly, were tested by means of auditorily presented and written items. Accuracy on auditory items was lower and response times were slower, indicating that the fleeting nature of the stimuli made the auditory modality more difficult (see also Kim & Godfroid, 2019). Moreover, we can expect declining auditory acuity in healthy ageing, which will compound the challenge for older learners. In keeping with this, we found an advantage for younger participants within the 13-year age range of our sample on the aptitude subtests assessing phonetic coding ability (LLAMA D and E).

Response times on the posttest revealed a somewhat unexpected pattern. First, participants in the incidental condition responded significantly faster than participants in the explicit condition but still achieved the same level of accuracy. Put differently, we can observe more efficient behaviour than in the explicit learners: despite responding more quickly, incidental participants performed just as well. Bearing in mind that the posttest was untimed and response times were leisurely across the board, we conjectured that faster responses were above all indicative of confidence in one's knowledge. It is possible to further conjecture that greater confidence may at least in part have arisen from more effortful and thus deeper processing encouraged by the absence of metalinguistic explanations and resulting attempts to independently work out the underlying systematicities of the target feature.

Taken together, several of our results point towards a role for confidence. Speed of response was not related to any of the aptitude measures. More importantly, the explicit condition may have encouraged participants to reflect carefully, thus leading to more cautious and therefore slower response behaviour. Having been presented with a metalinguistic description of the target feature, participants in the explicit group would have been aware of the full complexity of the inflectional paradigm that constituted the learning target and may thus have been more likely to doubt themselves and their knowledge. Indeed, correlations within groups showed that in the explicit group, faster responses were associated with greater accuracy. All in all, then, in the present study, slower responses appear to be indicative of lengthy reflection, hesitation, and self-doubt.

While admittedly speculative to some extent, this line of argument sheds new light on claims about explicit vs. implicit learning in older adults. Only recently, Cox and Sanz (2023) proposed that time constraints affected explicit but seemingly not implicit learning, and that if time pressure were removed in explicit learning contexts, older adults would be better able to bring to bear their explicit knowledge. Our findings do not support this hypothesis, since longer response times following learning in an explicit condition resulted in less accurate responses. Needless to say, this result is based on a specific learning context as operationalized in a single study and can therefore not be generalized; further investigation is clearly warranted.

2 Language learning aptitude in older adults' additional language learning

The present study has confirmed that the LLAMA suite and the SRT task are viable aptitude measures for use with older adults, thus substantiating the findings from the only previous study employing both of these measures with learners aged 60+ (Roehr-Brackin et al., 2023). As discussed in the preceding section, several LLAMA subtests proved to be significant predictors of participants' learning of an L2 at beginner level, supporting the findings of the only other available study to date that has made use of the LLAMA test battery as a predictor of late-life L2 learning (Donnerer & Roehr-Brackin, in press).

Our factor analysis has replicated the previously reported finding that LLAMA B (associative memory), E (sound-symbol association), and F (language-analytic ability) constitute measures of explicit aptitude, and LLAMA D (auditory pattern recognition) and the SRT task potentially constitute measures of implicit aptitude (Roehr-Brackin et al., 2023). These results provide a more unequivocal picture than the diverse results obtained with younger samples. In the latter case, LLAMA D either did (Granena, 2013) or did not (Roehr-Brackin et al., 2024) load separately from LLAMA B, E or F, and it either did (Pavlekovic & Roehr-Brackin, 2024) or did not (Roehr-Brackin et al., 2024) load together with the SRT task. While further evidence clearly needs to be sought, the fully convergent findings of the two studies involving late-life learners could suggest that LLAMA D works better as a measure of implicit aptitude in older than it does in younger adults.

Such a situation is certainly feasible if we conceptualize implicit aptitude as a cognitive proclivity rather than as a context-independent capacity (for discussion of this argument and LLAMA D, see Iizuka & DeKeyser, 2024; Pavlekovic & Roehr-Brackin, 2024). In particular, it sits well with two of our findings that were also observed in a previous study using the same measures (Roehr-Brackin et al., 2023): first, participants who were still working outperformed retired participants in terms of implicit aptitude; second, LLAMA D performance was associated with a more positive self-concept. It is possible to argue that being part of the workforce keeps individuals on their toes, requiring them to draw on implicit abilities on a regular basis and thus potentially boosting their confidence. The self-concept scale asked participants to (dis)agree with statements about their general health, happiness, and verbal, cognitive, and manual abilities. No doubt, both (meta)cognitive and affective factors come into play here. How do people feel within and about themselves? And how confident are they with regard to their own abilities? Viewed in this light, it is perhaps unsurprising that a more positive self-concept is associated with better performance on a measure of implicit aptitude where keeping one's nerve and going with the flow may well be the best route to success.

VII Conclusion

In conclusion, the present study has demonstrated that the LLAMA aptitude suite can predict older adults' learning of the beginnings of a new language. Moreover, our findings suggest that whether we encourage explicit or incidental learning with our instructional approach may be less relevant for third-age learners than the role of input modality

(auditory vs. written). Written L2 input appears to be most accessible to (educated, literate) late-life learners, whereas auditory input is altogether more challenging for such a population. Finally, participants' self-concept, including their confidence in themselves, their learning, and their knowledge, has emerged as an important factor, linking the (meta)cognitive and the socioaffective domain.

It goes without saying that the present study had its limitations. First, it would have been preferable if the entire participant sample had been recruited simultaneously, so fully matched experimental groups could have been created. Second, and related to the previous point, aptitude testing would have ideally taken place prior to presentation of the instructional materials in order to allocate participants to experimental groups based on a matched ranking. Third, the duration of the experimental treatment was short, so that any conclusions we can draw from the current findings are limited to the earliest stages of additional language learning.

Future studies should seek to further the research agenda by investigating the learning of other target features and languages, and by including productive skills over and above receptive skills in both learning materials and outcome measures. Taking forward the specific implications of the present study, a comparison of the effects of different input modalities (written or auditory) under different learning conditions (explicit or incidental) in a fully crossed design would appear to be the next logical step.

Last but not least, it is important to move beyond socioeconomically privileged samples in order to arrive at truly generalizable findings. Collectively, we have taken the step from WEIRD, participants from western, educated, industrialized, rich democracies (Andringa & Godfroid, 2020), to WEIRDO (WEIRD + old), but should now attempt to meet the challenge of including less privileged late-life learners.

Author note

The anonymized data used in this study will be made available after publication of the paper.

Acknowledgments

The research presented in this paper was funded by the British Academy/Leverhulme Trust (grant reference SRG23\230787), whose support is gratefully acknowledged. We also thank our participants for their willingness to engage with the study.

Funding

The authors disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: The research reported in this paper was funded by the British Academy/Leverhulme Trust, award reference SRG23\230787.

Ethical approval

Ethical approval for the study was granted by the University of Essex, Faculty of Social Sciences Ethics Subcommittee 3 (reference ETH2324-0386). Participants provided informed written consent.

ORCID iDs

Karen Roehr-Brackin  <https://orcid.org/0000-0002-1152-5852>

Renato Pavlekovic  <https://orcid.org/0000-0002-5955-0379>

Note

1. A reviewer questioned the use of the label “incidental.” We opted for this term because learners were not directly prompted to focus on form. Therefore, the condition in question did not meet the criteria of inductive instruction as per Norris and Ortega’s (2001) definition. As discussed in the background section, we cannot (and do not) make any claims about the learning processes participants actually engaged in.

References

- Andringa, S., & Godfroid, A. (2020). Sampling bias and the problem of generalizability in applied linguistics. *Annual Review of Applied Linguistics*, 40, 134–142. <https://doi.org/10.1017/S0267190520000033>
- Antoniou, M., Gunasekera, G.M., & Wong, P.C.M. (2013). Foreign language training as cognitive therapy for age-related cognitive decline: A hypothesis for future research. *Neuroscience and Biobehavioral Reviews*, 37(10), 2689–2698. <https://doi.org/10.1016/j.neubiorev.2013.09.004>
- Cox, J.G. (2017). Explicit instruction, bilingualism, and the older adult learner. *Studies in Second Language Acquisition*, 39(1), 29–58. <https://doi.org/10.1017/S0272263115000364>
- Cox, J.G., & Sanz, C. (2015). Deconstructing PI for the ages: Explicit instruction vs. practice in young and older adult bilinguals. *IRAL*, 53(2), 225–248. <https://doi.org/10.1515/iral-2015-0011>
- Cox, J.G., & Sanz, C. (2023). Learning a new language in older age: Psycholinguistic and cross-disciplinary approaches. In A. Godfroid & H. Hopp (Eds.), *The Routledge handbook of second language acquisition and psycholinguistics* (pp. 85–96). Routledge. <https://doi.org/10.4324/9781003018872-9>
- Donnerer, H., & Roehr-Brackin, K. (in press). Learning an additional language in the third age: Comparing a monolingual and a multilingual approach. In J.E. Wainwright & U. Stickler (Eds.), *Language learning later in life*. University of Toronto Press.
- Dörnyei, Z. (2009). *The psychology of second language acquisition*. Oxford University Press.
- Ellis, N.C. (1994). Consciousness in second language learning: Psychological perspectives on the role of conscious processes in vocabulary acquisition. *AILA Review*, 11, 37–56.
- Ellis, R. (2004). The definition and measurement of L2 explicit knowledge. *Language Learning*, 54(2), 227–275. <https://doi.org/10.1111/j.1467-9922.2004.00255.x>
- Gabrys-Barker, D. (2018). Introduction: The background. In D. Gabrys-Barker (Ed.), *Third age learners of foreign languages* (pp. xiii–xxvii). Multilingual Matters.
- Goo, J., Granena, G., Yilmaz, Y., & Novella, M. (2015). Implicit and explicit instruction in L2 learning: Norris & Ortega (2000) revisited and updated. In P. Rebuschat (Ed.), *Implicit and explicit learning of languages* (pp. 443–482). John Benjamins.
- Granena, G. (2013). Cognitive aptitudes for second language learning and the LLAMA language aptitude test. In G. Granena & M.H. Long (Eds.), *Sensitive periods, language aptitude, and ultimate L2 attainment* (pp. 105–129). John Benjamins.
- Granena, G. (2020). *Implicit language aptitude*. Cambridge University Press. <https://doi.org/10.1017/9781108625616>
- Hulstijn, J.H. (2005). Theoretical and empirical issues in the study of implicit and explicit second-language learning: Introduction. *Studies in Second Language Acquisition*, 27(2), 129–140.

- Iizuka, T., & DeKeyser, R. (2024). Scrutinizing LLAMA D as a measure of implicit learning aptitude. *Studies in Second Language Acquisition*, 46(1), 28–50. <https://doi.org/10.1017/S0272263122000559>
- Kaufman, S.B., DeYoung, C.G., Gray, J.R., Jiménez, L., Brown, J., & Mackintosh, N. (2010). Implicit learning as an ability. *Cognition*, 116(3), 321–340. <https://doi.org/10.1016/j.cognition.2010.05.011>
- Kim, K.M., & Godfroid, A. (2019). Should we listen or read? Modality effects in implicit and explicit knowledge. *The Modern Language Journal*, 103(3), 648–664. <https://doi.org/10.1111/modl.12583>
- Kliesch, M., Giroud, N., Pfenninger, S.E., & Meyer, M. (2018). Research on second language acquisition in old adulthood: What we have and what we need. In D. Gabryś-Barker (Ed.), *Third age learners of foreign languages* (pp. 48–73). Multilingual Matters.
- Kliesch, M., & Pfenninger, S.E. (2021). Cognitive and socioaffective predictors of L2 microdevelopment in late adulthood: A longitudinal intervention study. *Modern Language Journal*, 105(1), 237–266. <https://doi.org/10.1111/modl.12696>
- Lenet, A.E., Sanz, C., Lado, B., Howard, J.H., & Howard, D.V. (2011). Aging, pedagogical conditions, and differential success in SLA: An empirical study. In C. Sanz & R.P. Leow (Eds.), *Implicit and explicit language learning: Conditions, processes, and knowledge in SLA and bilingualism* (pp. 73–84). Georgetown University Press.
- Li, S. (2015). The associations between language aptitude and second language grammar acquisition: A meta-analytic review of five decades of research. *Applied Linguistics*, 36(3), 385–408. <https://doi.org/10.1093/applin/amu054>
- Li, S. (2016). The construct validity of language aptitude: A meta-analysis. *Studies in Second Language Acquisition*, 38(4), 801–842. <https://doi.org/10.1017/S027226311500042X>
- Li, S. (2022). Explicit and implicit language aptitudes. In S. Li, P. Hiver, & M. Papi (Eds.), *Routledge handbook of second language acquisition and individual differences* (pp. 37–53). Routledge. <https://doi.org/10.4324/9781003270546-4>
- Li, S., & DeKeyser, R. (2021). Implicit language aptitude: Conceptualizing the construct, validating the measures, and examining the evidence. *Studies in Second Language Acquisition*, 43(S3), 473–497. <https://doi.org/10.1017/S0272263121000024>
- Lindenberger, U. (2014). Human cognitive aging: Corriger la fortune? *Science*, 346(6209), 572–578. <https://doi.org/10.1126/science.1254403>
- Mackey, A., & Sachs, R. (2012). Older learners in SLA research: A first look at working memory, feedback, and L2 development. *Language Learning*, 62(3), 704–740. <https://doi.org/10.1111/j.1467-9922.2011.00649.x>
- Meara, P. (2005). *LLAMA language aptitude tests*. Lognostics.
- Meara, P., & Rogers, V. (2019). *The LLAMA tests v.3*. Lognostics.
- Norris, J.M., & Ortega, L. (2001). Does type of instruction make a difference? Substantive findings from a meta-analytic review. *Language Learning*, 51(1), 157–213. <https://doi.org/10.1111/j.1467-1770.2001.tb00017.x>
- Oxford, R. (2018). A developmental perspective on third-age learning. In D. Gabryś-Barker (Ed.), *Third age learners of foreign languages* (pp. 3–18). Multilingual Matters.
- Pavlekovic, R., & Roehr-Brackin, K. (2024). Aptitude for explicit and implicit learning. *Journal of the European Second Language Association*, 8(1), 1–17. <https://doi.org/10.22599/jesla.117>
- Pfenninger, S.E., & Kliesch, M. (2023). Variability as functional marker of second language development in older adult learners. *Studies in Second Language Acquisition*, 45(4), 1004–1030. <https://doi.org/10.1017/S0272263123000013>
- Pfenninger, S.E., & Polz, S. (2018). Foreign language learning in the third age: A pilot feasibility study on cognitive, socio-affective and linguistic drivers and benefits in relation to previous

- bilingualism of the learner. *Journal of the European Second Language Association*, 2(1), 1–13. <https://doi.org/10.22599/jesla.36>
- Piechurska-Kuciel, E., & Szyszka, M. (2018). Compensatory strategies in senior foreign language students. In D. Gabrys-Barker (Ed.), *Third age learners of foreign languages* (pp. 108–124). Multilingual Matters.
- Plonsky, L., & Oswald, F.L. (2014). How big is ‘big’? Interpreting effect sizes in L2 research. *Language Learning*, 64(4), 878–912. <https://doi.org/10.1111/lang.12079>
- Ramírez Gómez, D. (2016). *Language teaching and the older adult: The significance of experience*. Multilingual Matters.
- Roehr-Brackin, K. (2023). Language learning in older adults: Interdisciplinary perspectives. *Language Teaching*, 56(2), 292–295. <https://doi.org/10.1017/S0261444822000325>
- Roehr-Brackin, K. (2024). Explicit and implicit knowledge and learning of an additional language: A research agenda. *Language Teaching*, 57(1), 68–86. <https://doi.org/10.1017/S026144482200026X>
- Roehr-Brackin, K., Baranowska, K., Pavlekovic, R., & Scheffler, P. (2024). The role of individual learner differences in explicit language instruction. *The Modern Language Journal*, 108(4), 815–845. <https://doi.org/10.1111/modl.12963>
- Roehr-Brackin, K., Loaiza, V., & Pavlekovic, R. (2023). Language learning aptitude in older adults. *Journal of the European Second Language Association*, 7(1), 1–15. <https://doi.org/10.22599/jesla.93>
- Rogers, V., Meara, P., & Rogers, B. (2023). Testing language aptitude: LLAMA evolution and refinement. In Z.E. Wen, P. Skehan, & R.L. Sparks (Eds.), *Language aptitude: Theory and practice* (pp. 47–72). Cambridge University Press.
- Salthouse, T.A. (2004). What and when of cognitive aging. *Current Directions in Psychological Science*, 13(4), 140–144.
- Singleton, D. (2018). Really late learners: Some research contexts and some practical hints. In D. Gabrys-Barker (Ed.), *Third age learners of foreign languages* (pp. 19–30). Multilingual Matters.
- Singleton, D., & Pfenninger, S.E. (2018). L2 acquisition in childhood, adulthood and old age: Misreported and under-researched dimensions of the age factor. *Journal of Second Language Studies*, 1(2), 254–275. <https://doi.org/10.1075/jsls.00003.sin>
- Singleton, D., & Ryan, L. (2004). *Language acquisition: The age factor* (2nd ed.). Multilingual Matters.
- van der Ploeg, M., Keijzer, M., & Lowie, W. (2023). Language pedagogies and late-life language learning proficiency. *IRAL, Online First View*, 1–38. <https://doi.org/10.1515/iral-2023-0079>
- Ware, C., Damnee, S., Djabelkhir, L., Cristancho, V., Wu, Y.-H., Benovici, J., Pino, M., & Rigaud, A.-S. (2017). Maintaining cognitive functioning in healthy seniors with a technology-based foreign language program: A pilot feasibility study. *Frontiers in Aging Neuroscience*, 9(42), 1–10. <https://doi.org/10.3389/fnagi.2017.00042>
- Wen, Z.E., Skehan, P., & Sparks, R.L. (Eds.). (2023). *Language aptitude: Theory and practice*. Cambridge University Press. <https://doi.org/10.1017/9781009076463>