DOI: 10.1111/modl.12963

ORIGINAL ARTICLE

The Modern Language Journal

The role of individual learner differences in explicit language instruction

Karen Roehr-Brackin[1] | **Karolina Baranowska**[2] | **Renato Pavlekovic**[1] | **Paweł Scheffler**[2]

[1]Department of Language and Linguistics, University of Essex, Colchester, UK

[2]Department of Applied English Linguistics and Language Teaching, Adam Mickiewicz University, Poznań, Poland

Correspondence
Karen Roehr-Brackin, University of Essex, Department of Language and Linguistics, Wivenhoe Park, Colchester CO4 3SQ, UK.
Email: kroehr@essex.ac.uk

Abstract

Aptitude–treatment interaction (ATI) research is of both theoretical and practical interest to second language (L2) learning, since it provides insights into the processes linking learner-internal individual difference factors and learner-external contextual variables including instructional approach—variables that jointly determine L2 outcomes. The present study employed a full range of aptitude measures mapped onto four explicit instructional conditions: auditory inductive, written inductive, mixed inductive, and mixed deductive. International volunteers ($N = 136$) completed online language lessons in beginners' Polish targeting two morphological features. Participants' phonetic and language-analytic abilities, level of multilingualism, and age predicted L2 achievement. A cluster analysis identified four learner profiles: high aptitude, low aptitude, memory oriented, and analytically oriented. Deductive instruction seemed to neutralise individual differences in aptitude, while ATI effects were observed in the single-modality conditions, with auditory input favouring high-aptitude learners and written input favouring high-aptitude, analytically oriented, and memory-oriented learners. We discuss the theoretical and practical import of these findings by highlighting the "capital" afforded by prior language learning experience, over and above the role of cognitive ability. In addition to the inductive–deductive contrast in explicit instruction, we emphasise the importance of input modality, which has hitherto been neglected in the field.

KEYWORDS
aptitude–treatment interactions, explicit instruction, individual differences, input modality, language learning aptitude, learner profiles

The role of individual differences in instructed additional or second language (L2) learning can be examined in the context of aptitude–treatment interaction (ATI) research. Research in this paradigm investigates interactions between language learners' (cognitive) abilities—aptitude—and specific instructional approaches—treatment—based on the assumption that different individuals will benefit to different extents from different kinds of language learning and teaching materials. While ATI research has a relatively long tradition in the field of L2 learning, recent empirical studies have primarily focused on a few selected cognitive capacities such as working memory (WM) or language-analytic ability. In contrast, studies with high ecological validity that draw on a full complement of aptitude measures are scarce; moreover, none of them is recent.

The present study examined ATIs in four instructional conditions with online learning materials mirroring activities used in real-world language courses. The experimental conditions were mapped onto aptitude components in terms of input modality (auditory vs. written) and memory and analytic demands (inductive vs. deductive). A comprehensive test battery comprising measures of aptitude for explicit and implicit learning as well as WM was employed, with a view to identifying the benefits that learners with specific profiles would derive from the different approaches.

BACKGROUND

The importance of ATI research with its simultaneous focus on learner-internal individual difference variables, learner-external variables in the sense of learning setting and instructional approach, and L2 outcomes is widely acknowledged (DeKeyser, 2012, 2021; Kempe & Brooks, 2011; Kidd et al., 2018). Indeed, the appreciation of ATI effects can be traced back to classic work on language learning aptitude. Carroll (1990), for instance, argued that an ability describes a relation between characteristics of individuals and characteristics of the task they perform with varying degrees of success, and he encouraged researchers to examine variations in the difficulty of cognitive tasks as a function of their characteristics. More recently, scholars have likewise pointed out that ATI effects are to be expected (Grañena, 2013c; Kormos, 2013; Wen et al., 2017).

Construct definitions: Aptitude and treatment

For the present study, aptitude is defined as comprising aptitude for explicit learning and aptitude for implicit learning, in accordance with current theorising (Grañena, 2013b, 2016, 2020). Aptitude for explicit learning (henceforth: explicit aptitude) includes the components of phonetic coding ability, language-analytic ability, and associative memory that are assessed in classic aptitude tests (Carroll, 1981, 1990; Skehan, 1998). Aptitude for implicit learning (henceforth: implicit aptitude) can be defined as "a cluster of cognitive abilities that (…) enables learners to conduct unconscious computation of the distributional and transitional probabilities of linguistic input" (Li & DeKeyser, 2021, p. 474).

Unlike explicit aptitude, implicit aptitude is a recent addition to the L2 learning research field and has therefore been investigated much less. It has been suggested that implicit aptitude comprises both domain-general cognitive abilities and domain-specific linguistic abilities. It is seen as a relatively stable predictor of achievement in terms of both learning rate and ultimate attainment in the L2. Over and above well-established evidence for variance in explicit learning, there is likewise some evidence for systematic variability in implicit learning, making implicit aptitude a relevant construct (Li & DeKeyser, 2021). At the same time, the tasks that have been used to measure implicit aptitude in L2 research to date are not necessarily highly reliable, often do not correlate with each other, and vary in their predictive power (Perruchet, 2021), so in this sense, caution is in order.

While both explicit and implicit aptitudes are componential constructs, they are distinct not only from each other (see Li, 2022, for an up-to-date overview) but also from WM. The argument that WM should be considered a component of aptitude because it is predictive of L2 attainment goes back to the 1990s (McLaughlin, 1995; Miyake & Friedman, 1998) and has since gained considerable support (Juffs & Harrington, 2011; Linck et al., 2014; Wen et al., 2017).

The notion of treatment refers to the characteristics of the instructional approach learners experience and the teaching and learning materials they engage with. Complementary to the distinction between explicit and implicit aptitude, explicit and implicit instruction can be distinguished. An instructional treatment can be considered explicit "if rule explanation comprised any part of the instruction (…) or if learners were directly asked to attend to particular forms and to try to arrive at metalinguistic generalizations of their own" (Norris & Ortega, 2001, p. 167). The first type of explicit instruction is typically referred to as deductive (rules are provided), while the latter is referred to as inductive (learners are encouraged to focus on form but must identify any rules by themselves). In the absence of either of these criteria, the treatment can be regarded as implicit.

Theoretical and practical significance of aptitude–treatment interaction research

ATIs are of theoretical import, since they can give insight into the processes that link learner-internal and learner-external variables and can thus contribute to our understanding of how language learning operates (DeKeyser, 2012). Two theoretical approaches to L2 learning, which are both situated in an information-processing paradigm, directly refer to ATI. First, the proposal of a small number of aptitude complexes that differentially relate to learning outcomes under different processing conditions suggests that patterns of abilities need to be matched to learning tasks (Robinson, 2001, 2007). Aptitude complexes for (a) focus on form, (b) incidental learning via oral content, (c) via written content, and (d) explicit rule learning are posited. In each complex, different sets of cognitive abilities are expected to interact dynamically with instructional conditions and situational contexts.

Second, attempts have been made to map aptitude components onto information-processing stages that characterise the L2 learning process (Skehan, 2002, 2016). For instance, phonetic coding ability and phonological WM may be particularly relevant at the earliest input-processing stage, executive WM may be of primary importance for the subsequent stage of noticing, language-analytic ability and WM are required for pattern identification and generalisation, and so forth.

Theoretical considerations are complemented by practical relevance: Understanding ATIs can help optimise L2 instruction because ATI research can show whether and why a particular teaching approach works better with some learners than others. In other words, a particular instructional approach draws on certain cognitive abilities. If these are present in the learner, the approach will work well; if they are not present, the approach will work less well—unless targeted remedial steps are taken. This line of reasoning suggests that L2 instruction can play to learners' strengths (DeKeyser, 2012, 2021; Robinson, 2005). However, it may also be able to compensate for learners' weaknesses by providing dedicated support in the form of scaffolding and extra practice (Ranta, 2008; Sawyer & Ranta, 2001; Skehan, 1998; Vatz et al., 2013).

To exemplify, if a learner has strong phonetic coding ability, they would likely benefit from ample auditory input, since this would play to one of their strengths. At the same time, if a learner has weak phonetic coding ability, they may benefit from extra listening exercises and explicit pronunciation instruction (Ranta, 2008), with activities broken down into manageable steps. By the same token, learners with strong language-analytic ability are likely to identify patterns and impose structure on input independently. Conversely, learners with weak language-analytic ability may benefit from the scaffolding provided by form-focused instruction, such as the presentation of pedagogical grammar rules (Sawyer & Ranta, 2001).

The role of aptitude at different levels of proficiency and for learning different linguistic features

In accordance with the theoretical argument that specific aptitude components may be more or less relevant at different information-processing stages, it has also been suggested that different aptitude components may predict success at different levels of L2 proficiency (Doughty, 2019; Li, 2022). At a general level, empirical evidence suggests that aptitude is a stronger predictor in beginners and/or at lower levels of L2 proficiency, compared with more advanced learners and/or higher levels of proficiency (Li, 2016, 2022). Recent research concerned with the role of WM has yielded mixed findings. On the one hand, the relationship between WM and proficiency was found to become stronger as L2 proficiency increased (Linck & Weiss, 2011), while on the other hand, the explanatory power of WM decreased as L2 proficiency increased (Serafini & Sanz, 2016). It has been argued that phonological WM, which is responsible for the storage and rehearsal of phonological material, is more important at lower levels. Conversely, executive WM, which includes executive functions such as selective attention, inhibition, and switching, is more important at higher levels of proficiency (Serafini, 2017; see also Li, 2023). At the same time, meta-analytic results suggest that executive WM is an overall better predictor of L2 achievement than phonological WM (Linck et al., 2014).

The differential role of aptitude (components) at the global level of L2 proficiency (Li, 2015, 2016) is complemented by a differential role of aptitude (components) for the learning and use of L2 structures of varying complexity (Grañena, 2013b). For instance, it has been reported that learners of L2 English relied to a greater extent on aptitude, and in particular language-analytic ability, when making grammaticality judgments of a difficult structure (passive voice) compared with an easy structure (past progressive). Overall, aptitude was found to be a facilitator for the difficult structure, but it had only a minimal effect on outcomes for the easy structure (Yalçın & Spada, 2016). Thus, it appears that aptitude is more strongly implicated in the learning of more difficult target features compared with easier features.

Learner types

Aptitude components have not only been associated with different levels of proficiency and L2 features of varying difficulty but have also been drawn on for the purpose of deriving learner profiles. Research in the 1980s that made use of the Modern Language Aptitude Test (MLAT; Carroll & Sapon, 1959) identified two main profiles: analytically oriented and memory-oriented learners (Skehan, 1986; Wesche, 1981). Several years later, Skehan (1998) conjectured that strong language-analytic ability may play a role up to a certain level of L2 proficiency, but that, ultimately, excellent memory ability allowing for the retention of large quantities of linguistic material was the most crucial aptitude component. He further proposed that unsuccessful learners were above all constrained by weaknesses in phonetic coding ability. Clearly, this early argument for the differential role of aptitude components at different levels of achievement foreshadowed Skehan's (2002, 2016) subsequent proposal for the mapping of specific cognitive capacities to the information-processing stages in L2 learning, as outlined above.

In a study involving 150 adolescent first language (L1) French learners of L2 English studying in an intensive English as a second language programme in Canada, Ranta (2002) administered several L2 proficiency measures as well as an L1 metalinguistic task involving error detection and correction aimed at assessing learners' language-analytic ability. A subsequent cluster analysis led to the identification of four learner types: Learners in Cluster 1 exhibited strong performance on all L2 measures as well as the metalinguistic task, while learners in Cluster 4 exhibited weak performance across all these measures. Cluster 2 learners were weak on the L2 tests and average on the metalinguistic task—a pattern that was interpreted as potentially reflecting analytically oriented learners, but who were at too early a stage of L2 development to be able to truly benefit from their language-analytic ability. Cluster 3 learners were average on measures of L2 vocabulary and listening and below average on L2 cloze and the L1 metalinguistic task. This pattern was interpreted as indicative of a memory orientation. Thus, two of the learner types identified in Ranta's study were comparable with the profiles that had been put forward previously, while the other two types reflected profiles that are either weak across the board or strong across the board. It is interesting to note that although these early studies suggest that cluster analysis is a potentially fruitful approach to understanding ATI effects and deriving practical implications, the method has not been employed in recent studies, possibly because researchers have focused on the measurement of a limited number of aptitude components.

Empirical studies of aptitude–treatment interactions

Indeed, many recent ATI studies have concentrated on specific issues, such as the role of WM in the context of different instructional treatments, the effects of the differential distribution of practice over time, or the interaction of WM and aptitude components with different types of corrective feedback (see, e.g., chapters in DeKeyser, 2021; Li et al., 2019; Suzuki & DeKeyser, 2017). By contrast, and despite the recognised importance of ATI research for both learning theory and instructional practice, empirical studies that both draw on a full range of aptitude measures and investigate ATIs in ecologically valid learning contexts are surprisingly scarce. Two frequently cited studies that meet these criteria are worth noting: Wesche (1981) and Erlam (2005).

Wesche's (1981) study was carried out with Canadian public service workers who were trained in either English or French. Following aptitude testing, the completion of learning style questionnaires, and interviews, learners were matched with one of three training methods: (a) the default audiovisual method, which took an inductive approach, focused on linguistic structures sequenced by difficulty, encouraged memorisation, and made use of auditorily presented dialogues and oral drills, (b) a deductive–analytical approach with explicit explanations of grammar and pronunciation and practice of all four skills, and (c) a functional–situational approach, which focused on specific communicative situations, was sequenced according to language uses, and incorporated speaking practice. The deductive–analytical approach was developed for learners with strong L1 skills and an analytic orientation (~20% of the sample), whereas the functional–situational approach was developed for learners with good memory abilities (~10% of the sample). Results are reported in broad strokes, indicating that learners in (b) displayed greater interest and motivation, less anxiety, and more positive attitudes towards the teaching method they experienced; they also achieved higher scores on three of the four L2 outcome measures. Learners in (c) showed the same tendencies, though this statement is based on anecdotal evidence. In summary, the findings from Wesche's (1981) study suggest that an instructional approach that capitalises on learners' strengths leads to more successful outcomes in terms of both learner satisfaction and L2 attainment.

Nearly 25 years later, Erlam (2005) worked with L1 English-speaking adolescents in New Zealand learning L2 French as part of their compulsory education. Targeting the teaching and learning of direct object pronouns, groups of learners were exposed to one of the following instructional approaches: (a) deductive instruction, which provided explicit rule explanation, form-focused activities, output practice, and corrective feedback, (b) inductive instruction with no explicit metalinguistic explanations but encouragement to focus on form and output practice, and (c) structured input instruction, which included metalinguistic information and aural and written input-based activities but no output practice, error identification activities, and corrective feedback. Participants were tested on phonetic coding ability, language-analytic ability, and WM. Results show that inductive instruction benefited learners with strong language-analytic ability, structured input-based instruction benefited learners with strong language-analytic ability and WM, and deductive instruction was of equal benefit to all learners. These findings imply that deductive instruction had a levelling effect—that is, it minimised individual differences in aptitude. In sum, Erlam's (2005) results suggest that specific instructional approaches cannot only play to learners' strengths, but can also be used to compensate for weaknesses. The notion of deductive instruction as levelling the playing field between learners of different abilities has been substantiated subsequently (Hwu & Sun, 2012; Hwu et al., 2014; Li et al., 2019; Sanz et al., 2016) and is complemented by the argument that a communicative classroom is unlikely to have a levelling effect (Ranta, 2002).

Taking into account recent developments in the theorising and measurement of aptitude which have led to a distinction between explicit and implicit aptitude, Grañena and Yilmaz (2018) provided a research synthesis of ATI studies that (a) included both an implicit and an explicit instructional condition, (b) measured learners' improvement, (c) investigated the relationship between a cognitive ability and instructional conditions, and (d) were published in refereed journals. Out of an initial pool of 48 ATI studies, only 9 met these selection criteria, and only 2 of the 9 studies measured implicit cognitive abilities. The researchers reported that seven studies found significant relationships between learning outcomes under explicit conditions and explicit cognitive abilities; the two studies that investigated implicit cognitive abilities yielded no relationships with outcomes under explicit conditions. This is interpreted as a dissociation between implicit and explicit cognitive abilities under explicit instructional conditions. In other words, explicit abilities facilitate learning under explicit conditions, but implicit abilities do not.

The two studies that included implicit aptitude measures led to a negative correlation between outcomes under implicit conditions and implicit cognitive abilities and a positive correlation between outcomes under implicit conditions and explicit cognitive abilities. This suggests that, contrary to expectation, implicit abilities do not facilitate learning under implicit instructional conditions. What is more, it is explicit abilities that provide an advantage (Grañena & Yilmaz, 2018). While surprising at first glance, this finding chimes with earlier theoretical arguments that (explicit) aptitude may be relevant in both explicit and implicit learning conditions (Robinson, 2005; Skehan, 2002), and that (explicit) aptitude may in fact be more important in naturalistic (i.e., implicit, or at least incidental) learning conditions where no scaffolding is available (Abrahamsson & Hyltenstam, 2008). Having said this, the reported pattern of results is based on just two studies, so caution is in order.

Summary and research issues

Research to date has shown that while aptitude is an important predictor of L2 outcomes, its precise role depends on learners' level of L2 proficiency, with greater effects expected at lower or beginner levels. Moreover, the role of aptitude appears to vary depending on the L2 features that are being targeted, with greater effects expected for the learning of more difficult structures. The role of specific aptitude components also varies, depending on the above factors as well as on the nature of the outcome measures used and the linguistic domain in focus.

Empirical studies investigating ATI effects suggest that explicit aptitude facilitates learning under explicit conditions, as expected. Implicit aptitude does not appear to facilitate learning under implicit conditions, but given the limited evidence available to date, this result cannot be regarded as conclusive, so the role of implicit aptitude in different instructional contexts requires further scrutiny. Existing research has further shown that instructional approaches that play to learners' strengths can lead to greater learner satisfaction and better learning outcomes. At the same time, and just as importantly, it also seems to be possible to compensate for learners' weaknesses, as suggested by the finding that deductive instruction can minimise the impact of differences in cognitive abilities.

When comparing different treatment conditions, researchers have manipulated the availability or otherwise of metalinguistic information about the targeted L2 feature, the availability and nature of corrective feedback, and the type and timing of practice activities. By contrast, and to the best of our knowledge, no ATI research to date has undertaken a principled comparison of different input modalities (auditory, written, or a combination of the two)—although Robinson's (2001, 2007) aptitude complexes feature a theoretical distinction between auditory and written input—and there is recent evidence for the importance of modality in L2 learning (Kim & Godfroid, 2019) and testing. Using auditory and written grammaticality judgment tasks, Grañena (2013c) reported that high-aptitude participants showed improved performance in the written test modality. By way of explanation, she argued that high-aptitude participants may have been able to bring to bear their superior language-analytic abilities in the untimed written test format, but could not draw on their explicit knowledge and engage in monitoring to the same extent in the paced auditory test format (Grañena, 2013c). While this explanation is plausible, the question arises as to what role phonetic coding ability might have played in the given context—an issue that was not addressed in the analysis presented, given that participants were grouped according to global aptitude scores.

To conclude, ATI studies that aim for ecological validity with their instructional treatments and also measure a full range of aptitude components are rare, and, importantly, none of them are recent. In other words, there have been no attempts yet to investigate the interaction of different treatment conditions with the full range of aptitude components as conceptualised in current theorising: explicit aptitude, implicit aptitude, and WM—a gap our study set out to fill.

THE PRESENT STUDY

We addressed three research questions:

- RQ1. Which individual difference factors predict adult participants' performance on selected L2 features following a set of online language lessons?
- RQ2. To what extent does participants' performance in different treatment conditions with (a) auditory input, (b) written input, and (c) a combination of the two depend on their learner profiles?
- RQ3. To what extent does participants' performance in inductive and deductive treatment conditions depend on their learner profiles?

We addressed these questions in a quasi-experimental study targeting two morphological features of Polish. Online learning materials comprising four language lessons were provided in four different experimental conditions: auditory inductive, written inductive, mixed inductive, and mixed deductive, followed by two posttests. Participants' explicit and implicit aptitude and WM were measured, and a background questionnaire was administered. Participants were adult volunteers who had no knowledge of Polish or other Slavic languages.

METHOD

This section provides a full overview of the methodology, including the design and focus of the instructional materials, measurement instruments used, data collection procedure, participants, and approach to data analysis.

Instructional treatment

The instructional treatment targeted two morphological features of Polish: adjective–noun gender agreement and the genitive of negation, as summarised in Table 1.

TABLE 1 Target features.

Lessons	Structure	Examples
1–2	Adjective–noun gender agreement	czerwona walizka (feminine, nominative) *red suitcase* czerwony zeszyt (masculine, nominative *red notebook* czerwone krzesło (neuter, nominative) *red chair*
3–4	Genitive of negation	Kupiłam walizkę (feminine, accusative) → Nie kupiłam walizki (feminine, genitive) *I bought a suitcase → I didn't buy a suitcase* Kupiłam zeszyt (masculine, accusative) → Nie kupiłam zeszytu (masculine, genitive) *I bought a notebook → I didn't buy a notebook* Kupiłam krzesło (neuter, accusative) → Nie kupiłam krzesła (neuter, genitive) *I bought a chair → I didn't buy a chair*

These features were chosen because they are frequent in the language and could therefore be embedded in ecologically valid learning materials that required minimal knowledge of vocabulary—criteria that were important, given that participants were complete beginners and had to learn a set of vocabulary items prior to proceeding to the language lessons, as detailed below.

We hypothesised that the genitive of negation would be more difficult than adjective–noun gender agreement. Learning difficulty of a linguistic feature depends on a number of variables, including frequency in the input, perceptual salience, communicative redundancy, the relative transparency of a form–meaning mapping, and formal complexity (DeKeyser, 2005). We focused on the latter variable when hypothesising a difference in difficulty in the two target features, since they differ in terms of the number of choices to be made (Hulstijn & de Graaff, 1994) and thus in the number of steps to be taken to arrive at the correct form (Housen et al., 2005; Spada & Tomita, 2010). Adjective–noun gender agreement requires the distinction of three morphemes in the nominative case, whereas the genitive of negation requires the distinction of six morphemes, three in the accusative case in positive sentences and three in the genitive case in negative sentences. Moreover, some of these morphemes are identical with the three morphemes featuring in nominative adjective–noun gender agreement, while others are different (see Table 1). Thus, we hypothesised that the genitive of negation would be more costly to process, which is expected to result in greater difficulty (Housen & Simoens, 2016).

The target features were incorporated into four language lessons that were delivered via the Moodle online learning platform. Lessons 1 and 2 focused on adjective–noun gender agreement, and Lessons 3 and 4 focused on the genitive of negation. The lessons were built around a storyline of a couple moving house (Lesson 1), going shopping (Lesson 2), organising a housewarming party (Lesson 3), and going on a business trip (Lesson 4). We developed learning materials that (a) drew on as small a set of vocabulary items as possible, (b) provided pictorial support as much as possible, (c) featured the targets as frequently as possible, and (d) embedded use of the targets in an entertaining communicative setting, thus reflecting the nature of language learning materials used in the real world, whether in a classroom or online.

Each language lesson consisted of a series of slides in H5P software comprising presentation and controlled practice materials. Participants could move freely through the slides, and view them in any order and as often as they wished within the overall time limit for the given lesson (see below for details on procedure). The presentation slides were based around simple dialogues between the two main characters, while the practice slides consisted of multiple-choice exercises focused on the target features. Immediate feedback was provided, and participants could have as many attempts at the practice exercises as they wished within the time limit of the lesson. All activities were receptive in nature, that is, participants were not required to speak or write in the L2.

The lessons were developed in four treatment conditions: auditory inductive, written inductive, mixed inductive, and mixed deductive. Each lesson in the mixed deductive condition started with an English-language metalinguistic explanation of the targeted feature. Conversely, the inductive conditions did not include any metalinguistic explanations, but instead prompted participants to pay attention to the endings of words. Participants in the inductive groups had additional practice slides (5 each for Lessons 1 and 2, 10 each for Lessons 3 and 4). Additional practice was provided to ensure that more tokens than types were available, thus giving inductive group participants the chance to work out the underlying systematicities. Deductive group participants were provided with metalinguistic explanations, which effectively provided a shortcut.

The instructional conditions differed in terms of input modality. As the group labels indicate, participants in the auditory inductive group had auditory input only, participants in the written inductive group had written input only, and participants in the mixed groups had mixed input, in the sense that Lessons 1 and 4 were in auditory format and Lessons 2 and 3 were in written format. In all conditions, following the English-language metalinguistic explanation (deductive) or prompt to focus on form (inductive), each lesson started with a slide setting the scene for the storyline. This slide was always in dual modality and provided the scene-setting information in both Polish and English. Subsequent slides were in L2 Polish only. Example presentation slides from Lesson 1 in the auditory and written modality are shown in Figures 1 and 2, respectively (for further examples from all lessons, see Online Supporting Information A).

Prior to engaging with the language lessons, participants went through two vocabulary learning sessions, one preceding Lessons 1 and 2 and another preceding Lessons 3 and 4. Participants were told that they needed to learn to recognise the vocabulary in speech and writing. The vocabulary featuring in the lessons was presented via H5P in both auditory and written format. Binary Polish–English word pairs were shown and spoken aloud by L1 speakers of Polish, alternating between a female and a male speaker. In addition, illustrations were used whenever possible for concrete objects, colours, and so on. The vocabulary learning phases comprised presentation and controlled practice, and participants could move freely through each vocabulary session within the overall time limit. Each vocabulary learning session was followed by a vocabulary test, which participants had to pass with a minimum of 90% accuracy to proceed to the language lessons. Participants had a maximum of three attempts at passing the vocabulary test; if they failed three times, they were excluded from the study.

Instruments

Learning of the target features was assessed by means of two posttests following Lesson 2 and Lesson 4, respectively. Posttest 1 comprised 15 four-way multiple-choice items on adjective–noun gender agreement, and Posttest 2 comprised 25 four-way multiple-choice items on the genitive of negation. The posttests were administered in the same modality as the language lessons participants had experienced, so items were presented auditorily, in written format, or in mixed format, as required.

Language learning aptitude was measured by means of the LLAMA aptitude test battery,[1] with the LLAMA D subtest following an adjusted version 2 (Meara, 2005) and the remaining three subtests based on version 3 of the test suite (Meara & Rogers, 2019). LLAMA B is a measure of associative memory, requiring participants to remember 20 pairings of new words (form) and images of unknown creatures (meaning) following a 2-minute learning phase. LLAMA D and LLAMA E target phonetic coding ability: LLAMA D is a test of sound recognition and comprised an exposure phase during which participants listened to 10 sound strings and a testing phase including 40 items, 10 from the exposure phase that appeared twice, mixed in with 20 previously unheard items. LLAMA E is a

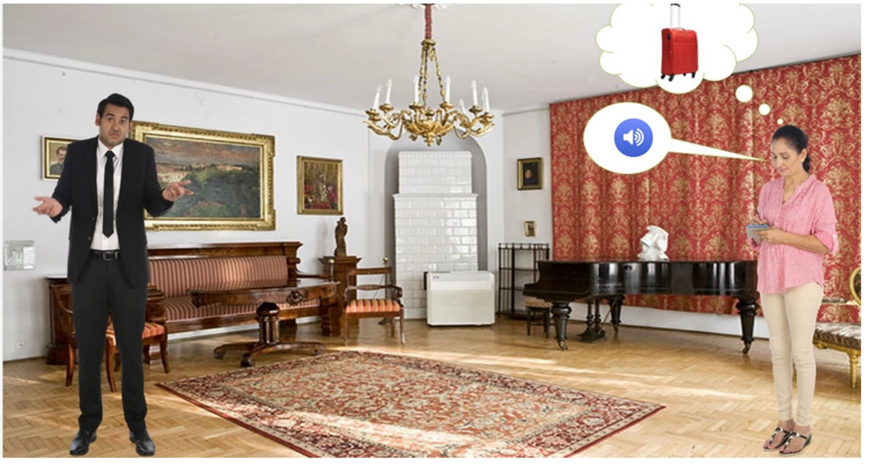

FIGURE 1 Example presentation slide from Lesson 1 (auditory modality). [Color figure can be viewed at wileyonlinelibrary.com]

FIGURE 2 Example presentation slide from Lesson 1 (written modality). [Color figure can be viewed at wileyonlinelibrary.com]

measure of sound–symbol association with a 2-minute learning phase during which participants are presented with pairings of 24 phonetic symbols and associated syllables. In the subsequent test phase, 20 two-syllable combinations must be matched with their associated phonetic symbols. LLAMA F tests language-analytic ability. During a 4-minute learning phase, participants are tasked with working out the rules of a mini artificial language by viewing sentences in that language and 20 images illustrating those sentences. The 20-item test phase as used in the present study required participants to construct sentences that accurately describe given images. Participants were allowed to take notes during the learning phases for LLAMA E and LLAMA F, if they wished. LLAMA B and LLAMA D were scored dichotomously, with incorrect answers on LLAMA D penalised to compensate for guessing; for LLAMA E and LLAMA F, we gave credit for partially correct answers. As this resulted in higher reliability than dichotomous scoring, the analyses reported below are based on partial-credit scores.

The LLAMA tests were complemented by a probabilistic serial reaction time (SRT) task (Kaufman et al., 2010) as a measure of implicit aptitude. This task requires participants to react to visual stimuli in the form of black squares that appear in one of the four possible locations on the computer screen. Participants are instructed to press a corresponding keyboard button as quickly and as accurately as possible. The sequence of stimuli is produced by a probabilistic rule; 85% of the time, the stimuli follow a training sequence, while the remaining 15% of the time, the stimuli follow a control sequence. A 60-trial practice phase preceded the task itself, which consisted of 8 blocks, each comprising 120 trials, resulting in a total of 960 trials. Implicit sequence learning ability is operationalised as the difference in mean response time between the control and training conditions.

WM was assessed by means of an operation span (O-Span) task (Unsworth et al., 2005). We selected a measure of executive WM, since previous research suggests that complex WM is an overall better predictor of L2 achievement than phonological WM (Linck et al., 2014), and since it has further been suggested that executive WM is a better predictor of the outcomes of explicit instruction and other instruction types that are cognitively demanding (Li, 2023; Williams, 2012)—criteria that apply to our experimental conditions.

In each trial on the O-Span task, participants first had to solve a simple mathematical problem and indicate whether a proposed solution shown on screen was correct or incorrect. They were then presented with a letter they needed to memorise. After a given sequence of mathematical problems followed by letters, participants selected the letters they could recall in order from an array shown on the computer screen. The task comprised 18 sets of letter sequences increasing in length from three to eight, totalling 99 letters. Data analysis is based on a partial-credit scoring system that resulted in higher reliability than dichotomous scoring.

All tests described above were programmed into PyschoPy (Peirce et al., 2019) and administered via the Pavlovia platform.

Following the lessons and tests, participants completed an exit questionnaire in which they provided basic demographic information, including age, gender, occupation, and level of education. As prior language learning experience and level of bi- or multilingualism have been shown to be associated with performance on cognitive ability measures (Cockcroft et al., 2019; Linck et al., 2014) and subsequent L2 achievement (Rogers et al., 2017), participants were asked to report previously learned languages and to indicate their level of proficiency in each of these languages on a simple 4-point scale using the descriptors “I know a few words and phrases, that’s all,” “I can have a simple conversation and/or understand basic information and/or read and write short and simple texts,” “I am a fairly confident user of the language, that is, I can speak, listen, read, and/or write it quite well,” and “I am a proficient user of the language.”

Furthermore, participants answered 12 items on a 5-point Likert scale and two open-ended questions that asked about their experience with the language learning materials. This allowed us to gauge participants’ perceptions, in line with the expectation that over and above cognitive abilities, attitudes can play a role in determining L2 success (Carroll, 1981). Finally, participants who had been allocated to an inductive condition were asked whether they thought they had been able to work out the rules underlying the targeted features. If they answered in the affirmative, they were prompted to spell out the rules, if they could. The exit questionnaire was administered via Qualtrics.

Piloting

The materials and measures were piloted to assess their quality and to try out the online set-up of the study. A total of 34 volunteers aged between 19 and 70 from mixed L1 backgrounds took part in the pilot study. The group included L1 speakers of English, Spanish, Chinese, Arabic, German, Thai, Turkish, and bilingual speakers with various combinations of L1 English, Spanish, Portuguese, or German with another L1. The technical side of the study worked as intended, and the various tests exhibited good or very good reliability. Therefore, all tests remained unchanged. Some items in the exit questionnaire were misunderstood by a minority of participants, so their wording was improved.

An exploratory correlation analysis revealed some associations between aptitude components and posttest performance in the inductive treatment groups despite extremely small sample sizes, suggesting that a larger study with sufficient statistical power to conduct inferential analyses may yield an ATI effect. Conversely, no correlations were found in the deductive group, indicating that the role of individual differences may have been neutralised, as reported in previous research.

TABLE 2 Timeline of the study.

Duration	Activities
Day 1	
30 minutes	Vocabulary learning
5 minutes	Vocabulary Test 1 – proceed if passed; go back to start if failed
35 minutes	Lesson 1
25 minutes	Lesson 2
10 minutes	Posttest 1 (adjective–noun gender agreement)
Day 2	
40 minutes	Vocabulary learning
5 minutes	Vocabulary Test 2 – proceed if passed; go back to start if failed
45 minutes	Lesson 3
35 minutes	Lesson 4
15 minutes	Posttest 2 (genitive of negation)
Day 3	
10 minutes	O-Span
30 minutes	LLAMA
30 minutes	SRT
15 minutes	Exit questionnaire

Abbreviations: O-Span, operation span task; SRT, serial reaction time task.

Procedure

Participation in the main study was spread over three separate days, and participants were asked to complete all three components within the same week. As all materials were provided online, they could otherwise work independently at times that were convenient to them, and they could take breaks as needed between lessons and tests as well as between LLAMA subtests and SRT blocks. Table 2 gives an overview of the timeline. The vocabulary learning sessions and language lessons had the exact time limits shown; timings for the tests and questionnaire are approximate.

Participation in the study was onerous and required a certain level of commitment from participants. Accordingly, recruitment of volunteers was a slow process that extended over several months. The criteria for participation were as follows: aged 18 or over, knowledge of English sufficient for the comprehension of instructions and the exit questionnaire, no knowledge of Polish or other Slavic languages. Participants were allocated to instructional conditions as they came forward, and each participant was provided with a personalised link that gave access to full information about the study and a consent form. Signing of the form triggered an automated chain of timed emails with further personalised links, allowing participants to move between learning and testing platforms in line with the progression of the study.

Participants

A total of 246 international volunteers came forward and were allocated to instructional conditions. Out of these, 176 participants actually started the study. The final sample comprised 136 participants who completed the exit questionnaire and thereby indicated their intention to finish the study. This data set is not complete, however, so participant numbers vary between treatment groups and between individual analyses, as reported below. Out of the 176 participants who began the study, 7 had to be excluded because they failed a vocabulary test three times; the other 33 decided not to continue to the exit questionnaire.

The final sample of 136 volunteers comprised 101 females, 32 males, 2 participants who identified as nonbinary, and 1 who preferred not to disclose their gender identity. Participants' ages ranged from 18 to 62 (M = 22.86, SD = 6.86). A total of 105 participants were full-time students, 10 were in full-time employment, and 21 were studying and/or working part time. Participants reported 14 different L1 backgrounds and 7 bi- or multilingual L1 combinations. The most frequently reported L1s were Chinese (n = 91), Japanese (n = 15), and Turkish (n = 6). On average, participants reported knowledge of 2.34 additional languages with a mean proficiency level of 2.27 on our self-assessment scale (min = 1, max = 4). In order to capture both quantity and quality of additional language knowledge, we calculated a multilingualism score for each participant by summing their z-scores for number of L2s and mean level of L2s. The final sample was distributed somewhat unevenly across the experimental conditions: auditory inductive, n = 33; written inductive, n = 45; mixed inductive, n = 31; mixed deductive, n = 27.

Data analysis

Reliability indices were computed for all instruments. Reliability of the posttests was good or very good, ranging from Cronbach's alpha = 0.73 to 0.86. The 12-item scale investigating participants' perceptions of the language lessons proved highly reliable: alpha = 0.91. Reliability of the LLAMA subtests based on version 3 was excellent, suggesting that recent improvements to the test suite have had the desired impact (Rogers et al., 2023): LLAMA B = 0.95, LLAMA E = 0.95, LLAMA F = 0.97. LLAMA D was based on version 2 and resulted in alpha = 0.63.

Reliability of the SRT task was calculated using Spearman–Brown split-half and resulted in an index of 0.52, which is either in line with or slightly higher than indices reported in previous research (Grañena, 2013b; Kaufman et al., 2010; Roehr-Brackin et al., 2023; Suzuki & DeKeyser, 2017). In general, reliability of above 0.4 is deemed acceptable for measures of implicit processes (Grañena, 2020). Participants who had 50% or more missing data were excluded from analyses involving the SRT task. The O-Span task proved to be highly reliable: alpha = 0.91. We scrutinised participants' performance on the mathematical problems that preceded letter recall. Mean accuracy was high at 93%, indicating that participants engaged with the task as intended. Two outliers who scored 50% accuracy or below were removed from analyses involving the O-Span task.

Normality of all variables both for the whole sample and by treatment condition was investigated by means of Shapiro–Wilk tests. Significant divergence from a normal distribution was observed in a number of instances, and analyses involving these variables employed nonparametric statistical tests, as detailed below. Detailed normality statistics and reliability indices can be found in the subsequent section.

To begin with, descriptive statistics were calculated, followed by inferential comparisons between groups and between posttests. Correlational analyses were run to examine the relationships between cognitive ability measures, background variables, and posttest scores. The number of cognitive variables was reduced by means of a factor analysis, and hierarchical regression analyses were employed to identify the predictors of L2 achievement in the sample as a whole (RQ1). We performed an agglomerative hierarchical cluster analysis to identify groups of learners sharing the same profile. In order to investigate ATI effects, we ran a robust ANOVA with bootstrapping followed by pairwise comparisons to examine whether learners with specific profiles derived differential benefits from specific instructional conditions (RQ2 and RQ3).

TABLE 3 Descriptive statistics for cognitive variables: Whole sample.

		M		*SD*						
Variable	***N***	**Raw**	**Cor.**	**Raw**	**Cor.**	**Min**	**Max**	**Skew**	**S–W (*p*)**	**α**
LLAMA B	134	8.76	43.81	7.00	35.00	0	20	0.413	.001	0.948
LLAMA D	134	10.03	25.07	8.91	22.28	0	28	0.463	.001	0.631
LLAMA E	134	17.04	42.59	13.29	33.23	1	40	0.429	.001	0.954
LLAMA F	134	56.54	47.12	36.84	30.70	8	119	0.169	.001	0.973
SRT	83	14.89		15.28		−33	54.41	0.136	.267	0.405
OSPAN	132	77.18	77.96	21.46	21.67	0	99	−1.370	.001	0.905

Note: Min and Max are raw values.
Abbreviations: Cor., corrected (percentage) score; OSPAN, operation span task; SRT, serial reaction time task; S–W, Shapiro–Wilk test of normality.

TABLE 4 Descriptive statistics for cognitive variables: Auditory condition.

		M		*SD*					
Variable	***N***	**Raw**	**Cor.**	**Raw**	**Cor.**	**Min**	**Max**	**Skew**	**S–W (*p*)**
LLAMA B	31	8.55	42.74	6.57	32.83	0	20	0.584	.008
LLAMA D	31	13.29	33.23	9.9	24.75	0	28	−0.097	.004
LLAMA E	31	20.06	50.16	14.8	37.00	1	40	0.088	.001
LLAMA F	31	66.36	55.30	37.05	30.88	8	117	−0.301	.006
SRT	22	15.83		17.34		−33.00	51.67	−0.709	.373
OSPAN	30	74.33	75.08	20.12	20.32	22	99	−0.848	.024

Note: Min and Max are raw values.
Abbreviations: Cor., corrected (percentage) score; OSPAN, operation span task; SRT, serial reaction time task; S–W, Shapiro–Wilk test of normality.

RESULTS

In the present study, we investigated which individual difference factors predicted participants' performance on two morphological features of Polish; the extent to which participants' performance in different treatment conditions with auditory input, written input, or a combination of the two depended on their profiles; and the extent to which participants' performance in inductive or deductive treatment conditions depended on their profiles. For an initial overview, Table 3 shows the descriptive statistics for the cognitive ability measures for the sample as a whole.

The descriptive statistics show a spread of scores on all variables, indicating individual learner differences across the board. It is worth noting that mean performance on the O-Span task is strong, whereas the mean LLAMA scores are all below 50%. LLAMA D was the most challenging subtest with a mean facility value of just 25%.

Tables 4–7 show the descriptive statistics for the cognitive ability measures by treatment group (for a graphic overview by treatment group, see boxplots in Online Supporting Information B).

Scrutiny of the tables suggests that the written inductive group performed best and the mixed deductive group worst—with the exception of the SRT task, where all groups performed very similarly. Inferential statistics (Kruskal–Wallis H) confirm significant differences between groups on LLAMA D ($p = .004$). Post hoc pairwise comparisons (one-way ANOVA) indicate that the written group (mean rank = 75.88) and the auditory group (mean rank = 80.47) outperformed the mixed deductive group (mean rank = 50.17)—that is, they had better sound recognition ability. If this conveyed an advantage for posttest performance, it is interesting to note that it seemingly applied to both the auditory and

TABLE 5 Descriptive statistics for cognitive variables: Written condition.

		M		*SD*					
Variable	***N***	**Raw**	**Cor.**	**Raw**	**Cor.**	**Min**	**Max**	**Skew**	**S–W (*p*)**
LLAMA B	45	8.20	41.00	6.96	34.78	0	20	0.505	.001
LLAMA D	45	11.96	29.89	9.11	22.78	0	28	0.180	.007
LLAMA E	45	19.04	47.61	13.14	32.84	1	40	0.094	.001
LLAMA F	45	60.49	50.41	37.46	31.22	8	119	0.039	.001
SRT	27	14.81		14.48		−13.33	53.33	0.337	.643
OSPAN	45	78.73	79.53	21.96	22.18	2	99	−1.630	.001

Note: Min and Max are raw values.
Abbreviations: Cor., corrected (percentage) score; OSPAN, operation span task; SRT, serial reaction time task; S–W, Shapiro–Wilk test of normality.

TABLE 6 Descriptive statistics for cognitive variables: Mixed inductive condition.

		M		*SD*					
Variable	***N***	**Raw**	**Cor.**	**Raw**	**Cor.**	**Min**	**Max**	**Skew**	**S–W (*p*)**
LLAMA B	31	10.19	50.97	7.33	36.66	0	20	0.154	.002
LLAMA D	31	7.48	18.71	7.21	18.03	0	24	0.717	.003
LLAMA E	31	15.32	38.31	12.48	31.21	2	39	0.675	.001
LLAMA F	31	52.40	43.67	35.27	29.39	9	116	0.334	.006
SRT	21	15.53		15.80		−5.01	54.41	1.110	.028
OSPAN	31	78.77	79.57	22.99	23.22	0	99	−1.752	.001

Note: Min and Max are raw values.
Abbreviations: Cor., corrected (percentage) score; OSPAN, operation span task; SRT, serial reaction time task; S–W, Shapiro–Wilk test of normality.

TABLE 7 Descriptive statistics for cognitive variables: Mixed deductive condition.

		M		*SD*					
Variable	***N***	**Raw**	**Cor.**	**Raw**	**Cor.**	**Min**	**Max**	**Skew**	**S–W (*p*)**
LLAMA B	27	8.30	41.48	7.33	36.63	0	20	0.466	.003
LLAMA D	27	6.00	15.00	7	17.49	0	24	1.288	.001
LLAMA E	27	12.18	30.46	11.54	28.86	2	40	1.331	.001
LLAMA F	27	43.46	36.22	34.75	28.96	8	113	0.813	.001
SRT	13	12.46		13.83		−10.89	30.98	−0.123	.437
OSPAN	26	76.54	77.31	20.85	21.06	24	99	−1.198	.004

Note: Min and Max are raw values.
Abbreviations: Cor., corrected (percentage) score; OSPAN, operation span task; SRT, serial reaction time task; S–W, Shapiro–Wilk test of normality.

the written condition, with the latter perhaps unexpected. There are no other statistical differences in cognitive ability between the groups.

Table 8 provides the descriptive statistics for the four treatment groups and the sample as a whole on Posttest 1 (adjective–noun gender agreement) and Posttest 2 (genitive of negation), respectively (for a graphic overview by treatment group, see boxplots in Online Supporting Information B).

TABLE 8 Descriptive statistics for Posttests 1 and 2.

		M		*SD*						
Group	***N***	**Raw**	**Cor.**	**Raw**	**Cor.**	**Min**	**Max**	**Skew**	**S–W (*p*)**	***α***
Posttest 1 (Adjective–noun gender agreement)										
Auditory	32	7	46.67	3.57	23.83	2	15	0.502	.102	0.763
Written	45	8.96	59.70	3.75	24.98	2	15	0.183	.026	0.786
Mixed I	31	6.06	40.43	3.35	22.31	1	13	0.454	.216	0.755
Mixed D	27	5.52	36.79	3.41	22.75	2	13	1.438	.001	0.755
Total	135	7.14	47.60	3.77	25.16	1	15	0.518	.001	
Posttest 2 (Genitive of negation)										
Auditory	32	11.16	44.63	5.26	21.05	3	25	1.214	.001	0.811
Written	45	14.22	56.89	5.83	23.32	6	25	0.347	.016	0.858
Mixed I	31	10.90	43.61	4.13	16.54	5	21	0.845	.041	0.728
Mixed D	27	8.96	35.85	3.96	15.83	4	20	1.398	.002	0.728
Total	135	11.68	46.73	5.32	21.28	3	25	0.881	.001	

Note: Min and Max are raw values.
Abbreviations: Cor., corrected (percentage) score; D, deductive; I, inductive; S–W, Shapiro–Wilk test of normality.

TABLE 9 Descriptive statistics for background variables: Whole sample.

Variable	***N***	***M***	***SD***	**Min**	**Max**	**Skew**	**S–W (*p*)**
Age	136	22.86	6.86	18	62	3.500	.001
Level of education	132	2.73	0.72	1	5	0.445	.001
Multilingualism score	116	0	1.36	−2.56	4.55	1.193	.001
Perceptions of instruction	136	3.56	0.72	1.33	5	−0.925	.001

Abbreviation: S–W, Shapiro–Wilk test of normality.

Table 8 indicates that across the sample as a whole, mean scores on the two posttests are virtually identical. None of the experimental groups exhibited statistical differences in performance between Posttest 1 and Posttest 2 either: Wilcoxon signed ranks, auditory, $p = .589$; written, $p = .439$; mixed inductive, $p = .367$; mixed deductive, $p = .770$.

A visual comparison across groups again suggests a strong performance by the written group and a weak performance by the mixed deductive group. Inferential statistics (Kruskal–Wallis H) confirm significant differences between groups on Posttest 1 ($p < .001$) and Posttest 2 ($p < .001$). Post hoc pairwise comparisons (one-way ANOVA) confirm the superior performance of the written group and the relatively weak performance of the mixed deductive group. Specifically, the written group (mean rank = 87.19) significantly outperformed the mixed inductive (mean rank 58.08) and mixed deductive group (mean rank = 48.46) on Posttest 1. The written group (mean rank = 85.61) further outperformed the mixed deductive group (mean rank = 46.39) on Posttest 2.

Table 9 provides an overview of the descriptive statistics for the background variables measured in the study.

With regard to the background variables, it is worth noting that participants had generally quite positive views of the instruction they experienced, with a mean of 3.6 on a 5-point scale. This pattern is in evidence in all treatment conditions as well, though least in the mixed deductive group (for descriptive statistics by treatment group, see Online Supporting Information C).

In order to identify relationships between variables, we ran correlations (Spearman's rho) between scores achieved on the posttests, the cognitive ability measures, and the background variables. The

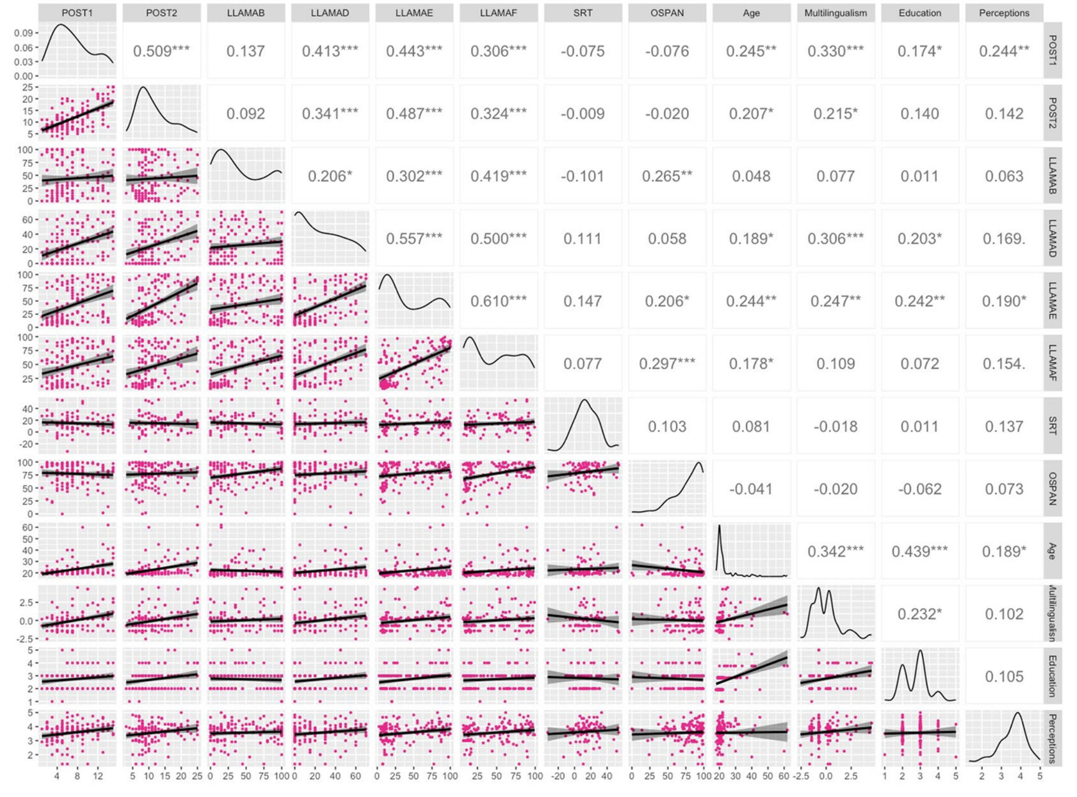

FIGURE 3 Correlations: Whole sample. OSPAN, operation span task; POST1, Posttest 1; POST2, Posttest 2; SRT, serial reaction time task. [Color figure can be viewed at wileyonlinelibrary.com]

results for the whole sample are shown in Figure 3. Results by treatment group are provided in Figures 4–7.

With regard to the correlational results, three main points are worth noting. First, the background variables of age and level of multilingualism correlate weakly to moderately with posttest performance in the sample as a whole. Note that the correlation with age is positive, that is, being older was associated with better posttest performance. Perceptions of the instructional materials and level of education likewise correlate weakly with Posttest 1, though the relationships lose their significance for Posttest 2. Second, Posttests 1 and 2 correlate, indicating consistent performance across the outcome measures for the two targeted features. Third, LLAMA D, LLAMA E, and LLAMA F consistently correlate with posttest performance. This pattern of results is largely borne out in the correlational analyses by treatment group. Specifically, scores on Posttest 1 and Posttest 2 correlate significantly in all groups, with the exception of the mixed inductive group where a trend in this direction was observed. Correlations with posttest scores of some or all the LLAMA subtests D, E, and F are in evidence in the inductive groups (written, auditory, and mixed), but not in the mixed deductive group, where none of the cognitive ability measures are significantly associated with posttest performance, although there is a trend towards a relationship between Posttest 1 and LLAMA D. Finally, the sample as a whole shows various associations between the LLAMA subtests and the O-Span task, though not with the SRT task.

In view of the correlational results, we conducted a factor analysis to reduce the number of variables. A principal component extraction with oblimin rotation (KMO = 0.70, Bartlett's test $p < .001$) resulted in three components with an eigenvalue (alpha) > 1 that explain 76% of the variance, as shown in Table 10.

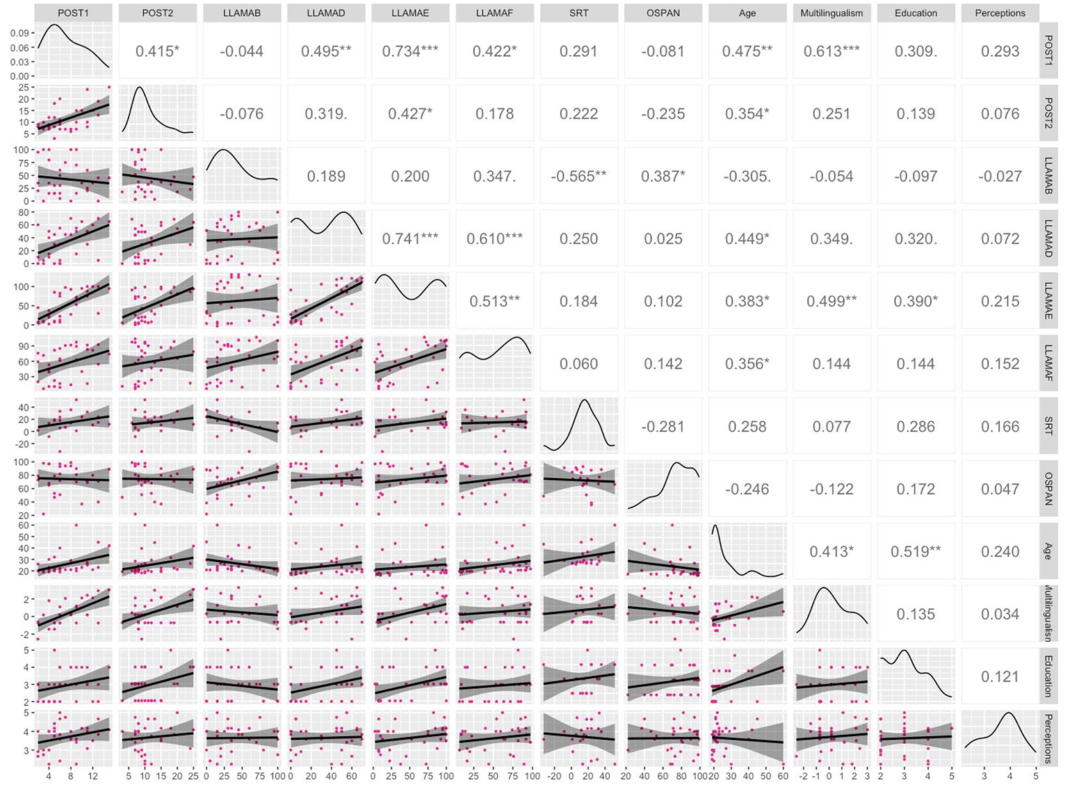

FIGURE 4 Correlations: Auditory group. OSPAN, operation span task; POST1, Posttest 1; POST2, Posttest 2; SRT, serial reaction time task. [Color figure can be viewed at wileyonlinelibrary.com]

The results in Table 10 show that LLAMA D, LLAMA E, and LLAMA F load on Factor 1 (alpha = 2.428), which captures phonetic and language-analytic abilities. LLAMA B and the O-Span task load on Factor 2 (alpha = 1.073), capturing memory ability. The SRT task as a measure of implicit sequence learning ability loads on Factor 3 (alpha = 1.067). These factor loadings are different from factor-analytic results reported in previous studies which included measures of both explicit and implicit aptitude. Specifically, LLAMA D has been found to load together with the SRT task and separately from LLAMA B, LLAMA E, and LLAMA F (Grañena, 2013a; Roehr-Brackin et al., 2023). The picture is complex, however, with a recent study revealing the same distribution of variables across factors, but the SRT task loading positively and LLAMA D loading negatively on the same factor (Pavlekovic & Roehr-Brackin, 2024).

In order to establish which individual difference factors predict participants' performance, we carried out two hierarchical regression analyses. The independent variables were the factor scores arising from the principal component analysis as well as the background variables of age, level of education, level of multilingualism, and perceptions of the instructional materials. Using the "enter" method, the independent variables were added into each regression model in descending order of correlational strength with the respective dependent variable: Posttest 1 as shown in Figure 8, and Posttest 2 as shown in Figure 9. The resulting models are shown in Tables 11 and 12.

Table 11 indicates that performance on Posttest 1 is significantly predicted by four variables explaining 24% of the total variance: phonetic and language-analytic abilities (8%), level of multilingualism (8%), age (5%), and perceptions of the instructional materials (3%). Table 12 shows that performance on Posttest 2 is significantly predicted by three variables accounting for 21% of the total variance:

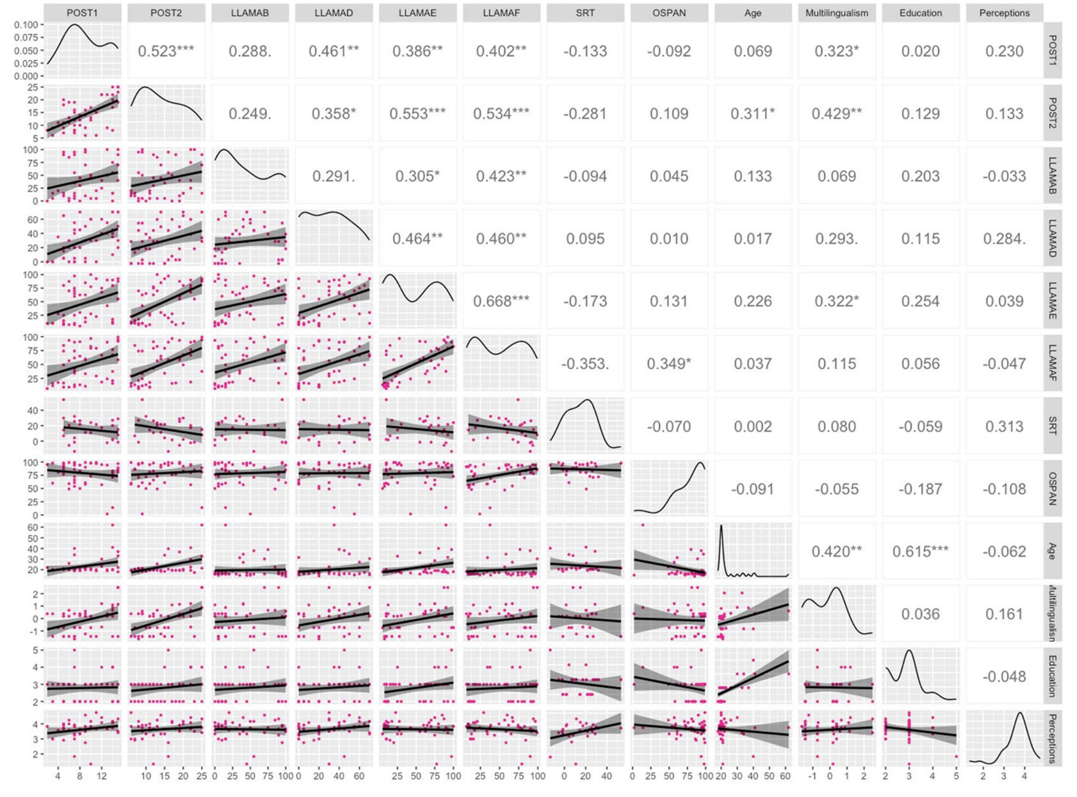

FIGURE 5 Correlations: Written group. OSPAN, operation span task; POST1, Posttest 1; POST2, Posttest 2; SRT, serial reaction time task. [Color figure can be viewed at wileyonlinelibrary.com]

phonetic and language-analytic abilities (10%), level of multilingualism (5%), and age (6%), with perceptions of the instructional materials no longer a significant predictor.

In order to investigate ATI effects, we performed two further analyses. First, we conducted an agglomerative hierarchical cluster analysis (following the steps in Staples & Biber, 2015) with a view to exploring whether our participants could be grouped according to specific profiles. The variables LLAMA B, LLAMA D, LLAMA E, LLAMA F, O-Span, level of multilingualism, and perceptions of the instructional materials were entered into the analysis. We included as many of the measured individual difference variables as possible in order to arrive at detailed learner profiles. Using Ward's method, we compared possible solutions based on the distance between fusion coefficients in the agglomeration schedule and settled on a 4-cluster solution for which all variables except perceptions of the instructional materials showed statistical differences between the clusters. The resulting learner profiles are shown in Table 13, while Figure 10 offers a visualisation in graphic format (see Online Supporting Information D for the agglomeration schedule, differences between clusters, and full descriptive statistics by cluster).

As Table 13 indicates, participants in Cluster 1 can be described as relatively inexperienced learners with generally low aptitude. By contrast, participants in Cluster 4 are highly multilingual learners with generally high aptitude. Learners in the remaining two clusters show more mixed profiles but are distinguishable in terms of whether they display strong memory ability (Cluster 2), or strong language-analytic ability (Cluster 3).

Second, and finally, we examined whether learners with different profiles as identified in the cluster analysis benefited to different extents from the four treatment conditions used in the present study. As it was our aim to identify treatment effects for the Polish learning experience as a whole, we

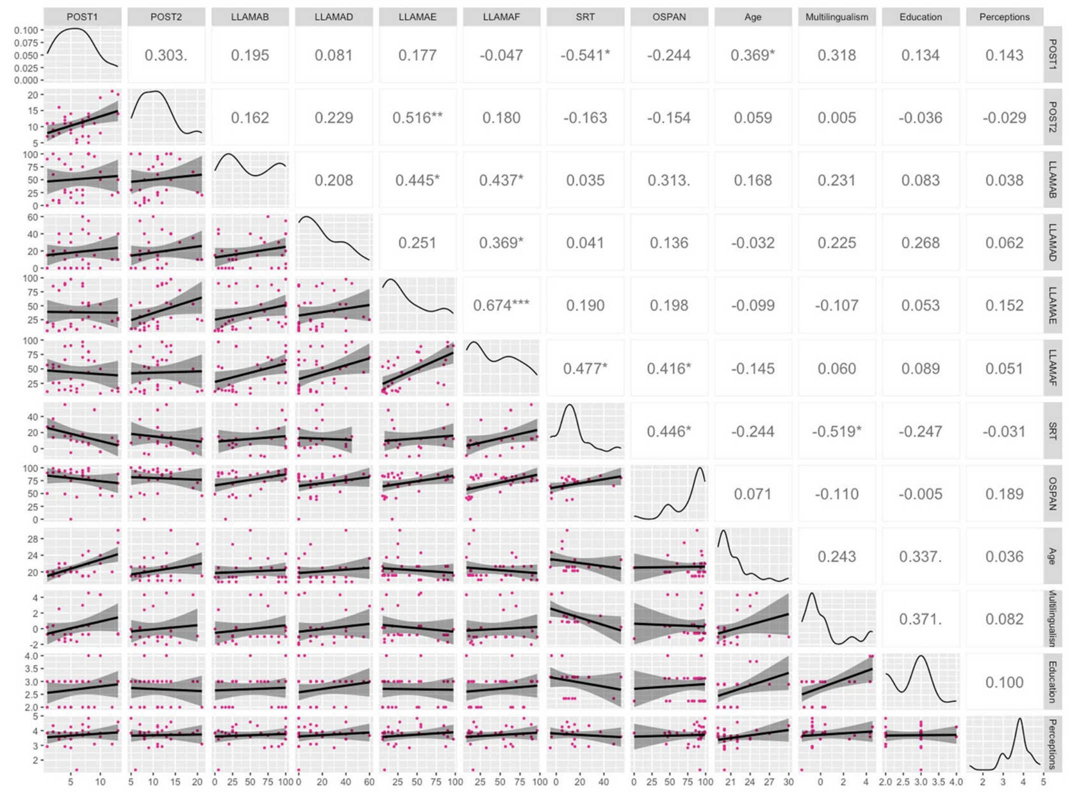

FIGURE 6 Correlations: Mixed inductive group. OSPAN, operation span task; POST1, Posttest 1; POST2, Posttest 2; SRT, serial reaction time task. [Color figure can be viewed at wileyonlinelibrary.com]

combined the mean scores from Posttest 1 and Posttest 2 into a mean posttest score as the outcome measure. Then, we conducted a nonparametric ANOVA with bootstrapping, entering treatment group and cluster membership as the between-subjects factors and combined posttest score as the dependent variable. The analysis yielded a nonsignificant main effect for cluster membership (p = .220) and a significant main effect for experimental group (p = .032). The Cluster × Treatment Group interaction was not significant (p = .488). Given the small N in some of the cells (see Table 14), the absence of statistical effects may be due to lack of power (Larson-Hall, 2016). A visual inspection of the posttest results achieved by participants with different profiles in the different instructional conditions as shown in Figure 11 suggests that in some of the experimental groups, performance differed substantially for learners in different clusters. Thus, following the steps outlined in Larson-Hall (2016), we ran pairwise comparisons (see Online Supporting Information E), which yielded statistical differences between clusters in the auditory and written groups. Specifically, high-aptitude learners outperformed low-aptitude learners in the auditory condition (p = .013). In the written condition, high-aptitude learners outperformed low-aptitude (p < .001) and memory-oriented learners (p = .010), and both analytically oriented (p < .001) and memory-oriented learners (p = .049) outperformed low-aptitude learners.

DISCUSSION

In what follows, the results arising from the present study are discussed in terms of predictors of L2 achievement at complete beginner level (RQ1) and observed ATI effects (RQ2 and RQ3).

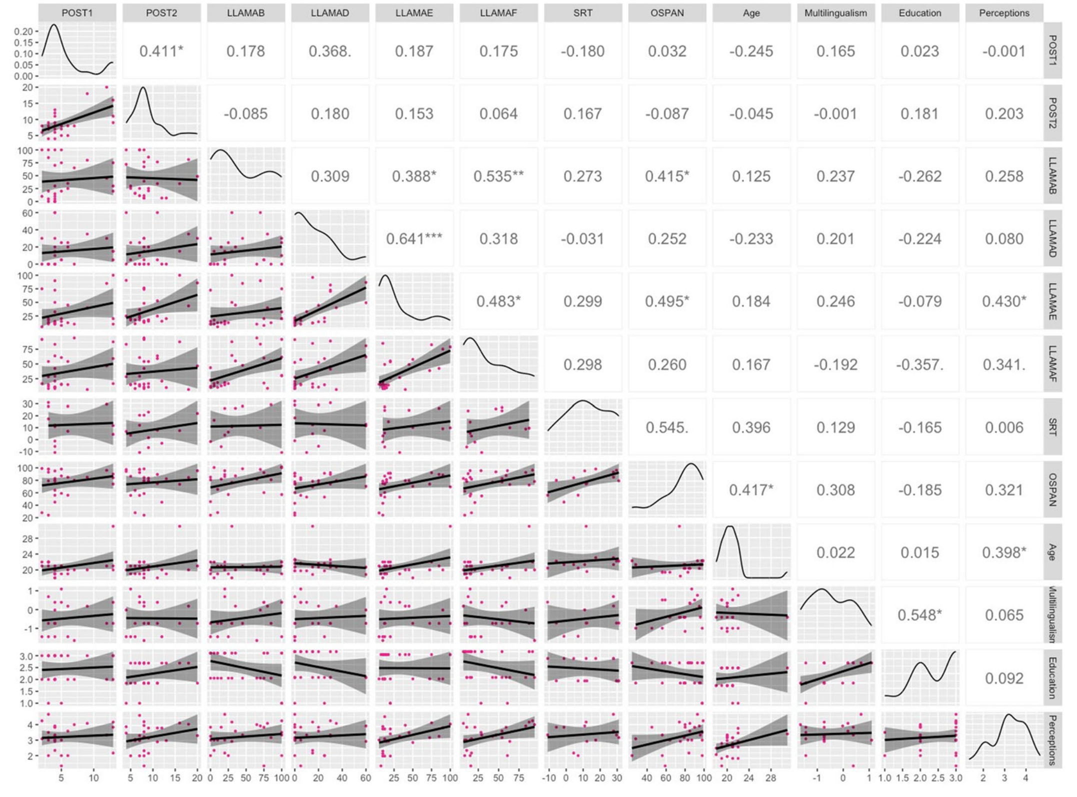

FIGURE 7 Correlations: Mixed deductive group. OSPAN, operation span task; POST1, Posttest 1; POST2, Posttest 2; SRT, serial reaction time task. [Color figure can be viewed at wileyonlinelibrary.com]

TABLE 10 Factor analysis structure matrix.

Factor	1	2	3	Communality (h^2)
LLAMA B	0.257	**0.803**	−0.255	0.776
LLAMA D	**0.839**	0.045	0.047	0.708
LLAMA E	**0.865**	0.238	0.120	0.819
LLAMA F	**0.791**	0.530	0.070	0.911
SRT	0.127	0.048	**0.927**	0.878
OSPAN	0.175	**0.765**	0.423	0.795

Note: The strongest loadings for each variable are indicated in bold. Abbreviations: OSPAN, operation span task; SRT, serial reaction time task.

Predictors of L2 achievement at beginner level

To begin with, it is worth noting that, contrary to expectation, our participants performed similarly on the posttests assessing the targeted features of adjective–noun gender agreement and genitive of negation. While the latter feature relied on a more extensive inflectional paradigm, this seemingly had no impact on learning difficulty. One possible reason for this is that the ordering of features was such that learners were scaffolded successfully from the less complex to the more complex paradigm and were therefore able to handle the increasing challenge. Another, potentially complementary reason

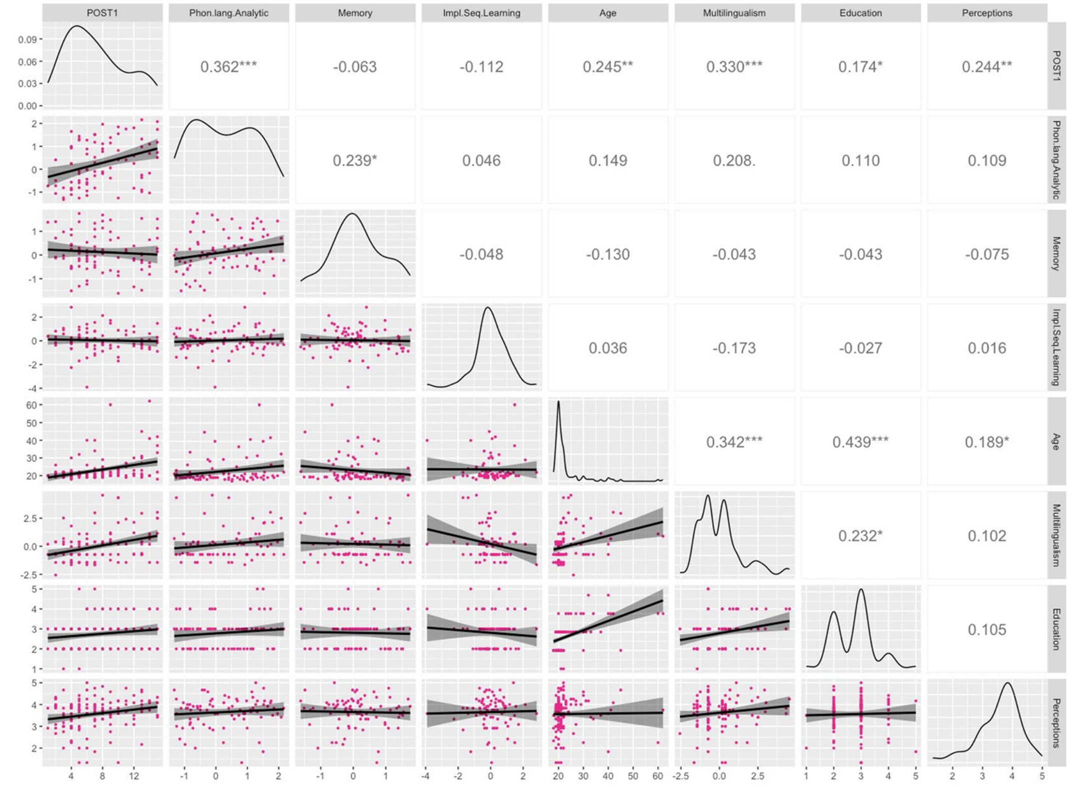

FIGURE 8 Correlations between predictor variables and Posttest 1: Whole sample. Impl.Seq.Learning, implicit sequence learning ability; Phon.Lang.Analytic, phonetic and language-analytic abilities; POST1, Posttest 1. [Color figure can be viewed at wileyonlinelibrary.com]

is the fact that the two features differed in quantitative terms (three vs. six inflections), whereas a qualitative difference in complexity may be required for learning difficulty to increase measurably.

Across the sample as a whole, we identified three consistent statistical predictors of L2 outcome in terms of posttest scores, with the phonetic and language-analytic ability factor emerging as the strongest predictor, followed by level of multilingualism and chronological age. Perceptions of the instructional materials explained a small amount of variance on Posttest 1, but this variable lost its significance for Posttest 2.

Taking the predictors in reverse order of strength, the minor role of learners' perceptions of the instructional materials may have lost significance over time as participants became accustomed to the nature of the instructional materials. What triggered stronger affective reactions in the first two lessons may have been less surprising in the last two lessons, so the small effect of this noncognitive variable could have been attenuated.

The role of participants' age may be surprising at first glance. Age correlated with level of education, but a partial correlation controlling for education confirmed that the role of age was not an artefact: Age was still associated with posttest scores even when level of education was factored out ($r_p = 0.338$, $p < .001$). A possible explanation for the role of age is that older participants put in more effort and engaged with the instructional materials with greater seriousness than younger participants. In other words, older participants behaved in a more mature manner. It should be borne in mind that in the context of the present study, older participants were mostly in their late 20s or in their 30s, and younger participants in their late teens or early 20s, so the terms are relative. There is no suggestion in our data that being 60 was more advantageous than being 30, for instance.

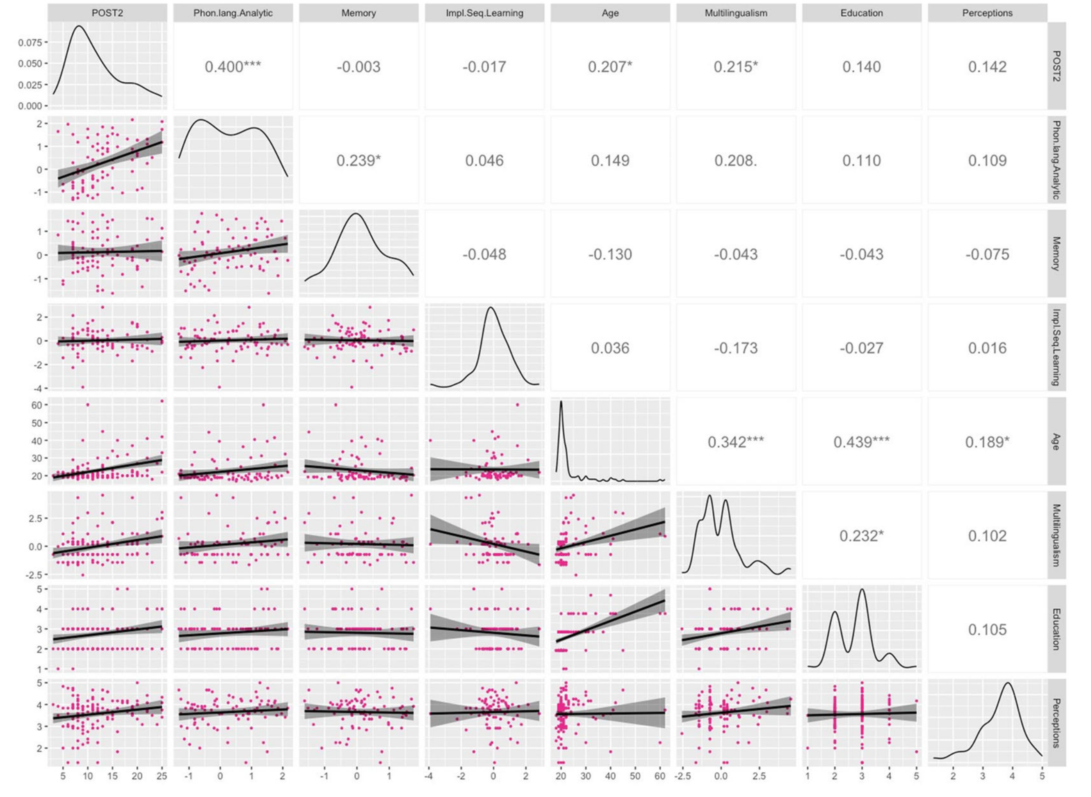

FIGURE 9 Correlations between predictor variables and Posttest 2: Whole sample. Impl.Seq.Learning, implicit sequence learning ability; Phon.Lang.Analytic, phonetic and language-analytic abilities; POST2, Posttest 2. [Color figure can be viewed at wileyonlinelibrary.com]

TABLE 11 Hierarchical regression for Posttest 1.

Variable	Model 1 B	Model 1 β	Model 2 B	Model 2 β	Model 3 B	Model 3 β	Model 4 B	Model 4 β
Phonetic and language-analytic abilities	1.353**	0.276	1.168**	0.238	0.996*	0.203	0.944*	0.193
Level of multilingualism			0.854**	0.285	0.642*	0.214	0.578*	0.193
Age					0.139**	0.255	0.143**	0.260
Perceptions of instructional materials							0.898*	0.171
R^2	0.076		0.156		0.214		0.243	
F	11.030**		12.300**		11.970**		10.50**	
ΔR^2	0.076		0.080		0.058		0.029	
ΔF	11.030 **		12.600**		9.720**		4.980*	

*$p < .05$; **$p < .001$.

The next predictor was level of multilingualism, that is, the quantity (number of languages) and quality (proficiency) of prior language learning. This result is in line with previous research, which has shown that language-learning experience can enhance the cognitive abilities facilitating the acquisition of subsequent languages (Cockcroft et al., 2019; Linck et al., 2014; Rogers et al., 2017)—a pattern also observable in the present sample. It is interesting to note that level of multilingualism has emerged

TABLE 12 Hierarchical regression for Posttest 2.

	Model 1		Model 2		Model 3	
Variable	**B**	**β**	**B**	**β**	**B**	**β**
Phonetic and language-analytic abilities	2.181**	0.315	1.975**	0.286	1.730**	0.250
Level of multilingualism			0.956**	0.227	0.654*	0.155
Age					0.198**	0.257
R^2	0.100		0.150		0.209	
F	14.810**		11.730**		11.610**	
ΔR^2	0.100		0.050		0.059	
ΔF	14.810 **		7.890**		9.820**	

*$p < .05$; **$p < .001$.

TABLE 13 Overview of learner profiles.

Cluster (*N*)	Multilingualism	Perceptions	LLAMA B	LLAMA D	LLAMA E	LLAMA F	OSPAN	Overall profile
1 (33)	Low	Positive	Very low	Very low	Very low	Very low	Medium	Low aptitude
2 (31)	Low–medium	Positive	High	Low	Low–medium	Low–medium	High	Memory oriented
3 (35)	Medium–high	Very positive	Low–medium	Medium	Medium	High	Medium	Analytically oriented
4 (15)	High	Very positive	High	High	Very high	Very high	High	High aptitude

Abbreviation: OSPAN, operation span task.

as a separate predictor from cognitive ability; experience built up over a period of time appears to be an asset in its own right, over and above the "raw material" of an individual's aptitude.

Phonetic and language-analytic ability was the strongest predictor of L2 outcome. The predictive power of aptitude is well evidenced (Li, 2016) and has been observed above all in explicit instructional conditions (Grañena & Yilmaz, 2018; Skehan, 2015)—conditions that also characterise the instructional treatments used in the present study. With regard to specific aptitude components, language-analytic ability has—unsurprisingly—been found to be important for grammar learning (Li, 2015), while phonetic coding ability can be expected to play a role at the earliest stages of acquisition (Skehan, 2016). Our results bear out these expectations: The participants were complete beginners, and the learning targets were two regular morphological features.

The memory factor and implicit sequence learning ability did not predict learners' posttest performance. The former finding is in line with the theoretical argument relating to the role of different aptitude components at different stages of acquisition, where memory is considered most important at more advanced stages when knowledge is proceduralised and automaticity of retrieval that manifests itself in fluent performance is achieved (Skehan, 2016). Clearly, this is a stage that our participants could not have reached after just four language lessons. At the same time, (working and associative) memory is also important for vocabulary acquisition. In the research design of the present study, this dimension was effectively neutralised because participants had met the required threshold of vocabulary-related memory prior to starting any language lessons.

Regarding the absence of predictive power of the SRT task, it has recently been proposed that implicit aptitude may not come into play until learners reach more advanced stages (Li, 2022), although evidence is still scarce, since very few studies to date have included measures of implicit cognitive abilities. Another possible reason may be found in the difference in stimulus domain between

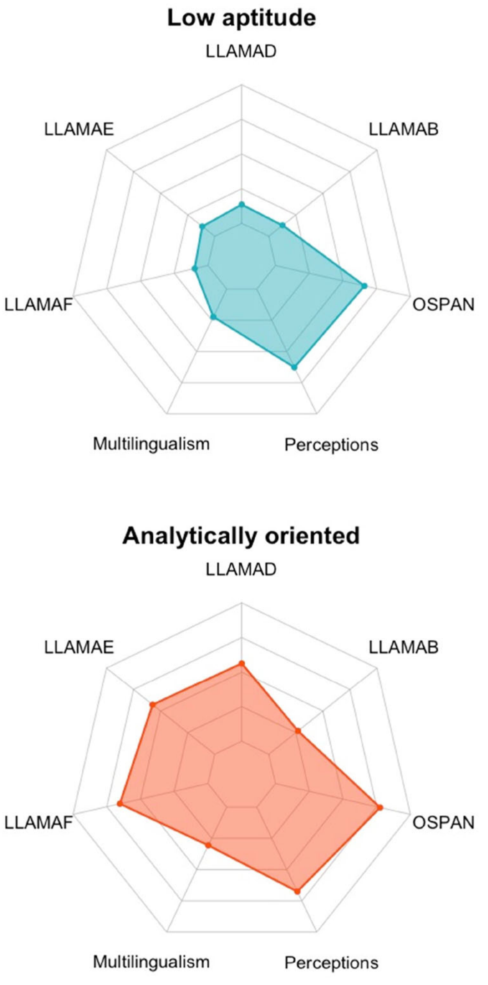

FIGURE 10 Radar plots of learner profiles. [Color figure can be viewed at wileyonlinelibrary.com]

the SRT task, which relies on nonverbal visual stimuli that are devoid of meaning, and the language learning task participants engaged in during the lessons. Beyond the nature of stimuli, the importance of sensory channel has come to the fore in our results relating to ATI effects, as discussed in the next section.

Aptitude–treatment interaction effects

In the present study, we compared four treatment conditions that differed by input modality and type of explicit instructional approach: auditory inductive, written inductive, mixed inductive, and mixed deductive. Although the mixed deductive group had the weakest posttest results, it was the only group in which none of the cognitive ability measures were associated with L2 outcome. This pattern of results is in keeping with the equalising effect of deductive instruction that has been reported in previous research (Erlam, 2005; Hwu & Sun, 2012; Hwu et al., 2014; Li et al., 2019; Sanz et al., 2016). Put differently, any impact of individual differences in aptitude may have been neutralised through the provision of instruction that included metalinguistic explanations of the targeted features at the

TABLE 14 Descriptive statistics for combined posttest scores by treatment condition and learner profile.

Treatment condition	Learner profile	Mean	*SD*	*N*
Auditory	Low aptitude	37.14	16.64	7
	Memory	41.11	16.52	3
	Analytic	50.53	16.67	10
	High aptitude	61.14	23.74	7
Written	Low aptitude	40.67	15.60	12
	Memory	56.75	18.41	8
	Analytic	67.07	17.60	15
	High aptitude	81.89	20.19	6
Mixed inductive	Low aptitude	38.48	20.88	7
	Memory	43.49	15.91	13
	Analytic	38.67	8.06	5
	High aptitude	44.67	.	1
Mixed deductive	Low aptitude	27.62	5.96	7
	Memory	45.33	24.36	7
	Analytic	43.07	18.39	5
	High aptitude	21.33	.	1

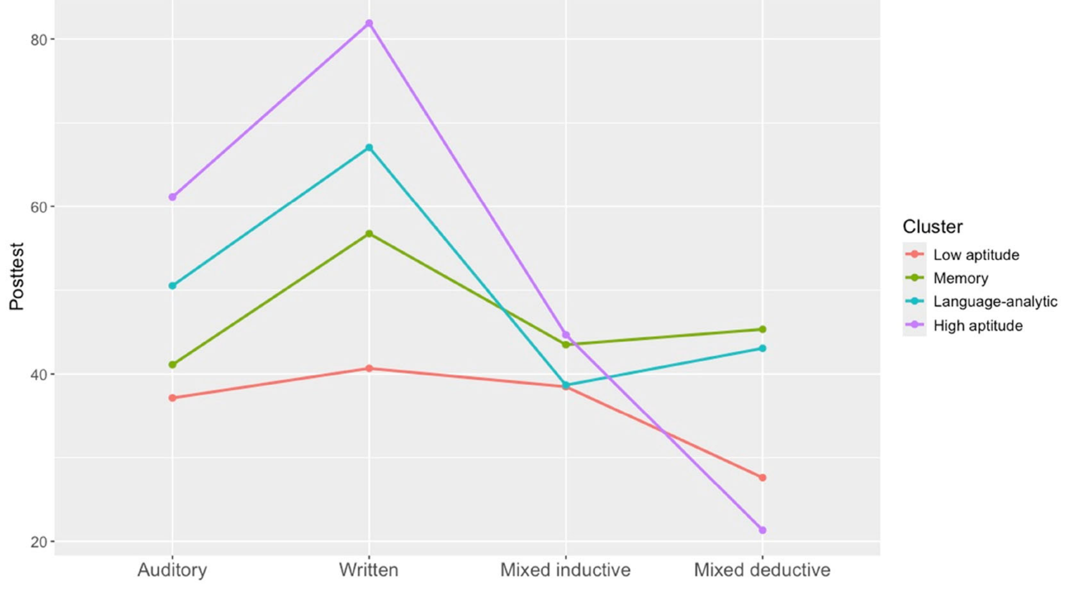

FIGURE 11 Aptitude–treatment interactions. [Color figure can be viewed at wileyonlinelibrary.com]

start of the presentation phase, thus scaffolding learners towards the target and thereby compensating for any weaknesses in language-analytic ability in particular. In addition, it is likely that the mixed-modality input compensated further, in the sense that learners could rely on their preferred input channel (auditory or visual), although this argument must remain speculative for now, as we shall see below.

Compared with the potential levelling effect of deductive instruction, empirical findings on learner profiles are not only rare, but also typically at least two decades old. Interestingly, our cluster analysis replicated an existing result (Ranta, 2002), despite the fact that we worked in very different circumstances and with a different set of instruments. While Ranta (2002) used various L2 measures as well as a metalinguistic task in L1 to assess language-analytic ability, the present study employed a full range of aptitude measures in accordance with current theorising. Ranta worked with a relatively homogeneous group of Canadian adolescents in an immersion classroom setting, while we worked with a highly heterogeneous group of adults who were exposed to a new language in the context of online lessons. Despite the fact that the methodological differences could hardly be more pronounced, both studies arrived at the same four learner types: low-aptitude individuals, high-aptitude individuals, memory-oriented learners, and analytically oriented learners. This constitutes powerful evidence for the psychological reality of memory-focused and language-analytic learner types, as suggested in early research (Skehan, 1986, 1998; Wesche, 1981), over and above global high-aptitude and low-aptitude profiles.

It is further worth noting that our learner profiles included an attitudinal component (perceptions of the instructional materials) and an experiential component (level of multilingualism). Whereas the former did not contribute significantly to the distinction between profiles, the latter did, thus reinforcing the finding arising from the regression analysis. In other words, prior language learning experience appears to be an asset in its own right, over and above the cognitive abilities that it is based on and that it might enhance. As individuals acquire additional languages and/or achieve higher levels of proficiency in the languages they know, they build up capital in the form of cumulative experience, which can pave the way to further learning.

While cognitive abilities are clearly critical for identifying meaningful learner profiles, the ATI effects we uncovered based on post hoc pairwise comparisons between clusters in the different experimental groups point to the importance of input modality and thus, potentially, also towards learners' perceptual abilities and/or preferences. We found ATI effects in the two treatment conditions that relied on a single input modality: In the auditory condition, high-aptitude learners were at an advantage. The high-aptitude participants were also the only learners who had high or very high phonetic coding ability (see Table 13), which no doubt helps explain this result. In the written condition, it was not only the high-aptitude learners who benefited, however: Analytically oriented and memory-oriented learners likewise outperformed low-aptitude participants. Therefore, in the written condition, higher levels of cognitive ability were needed as well, but these abilities were not (or were less) dependent on phonetic coding ability. Put differently, as a single input modality, the written condition brought individual differences to the fore, but it was more forgiving than the auditory condition because it allowed several learner types to succeed, since strengths in both language-analytic and memory ability could be exploited (see also Grañena, 2013c). Our finding therefore tallies with the argument that the more permanent nature of written stimuli can support (literate and educated) L2 learners at the beginner level (Kim & Godfroid, 2019). However, it appears that the advantage only takes effect if learners are strong in terms of language analysis and/or memory. For low-aptitude learners, input in a single modality seems to be unhelpful, regardless of whether it is auditory or written. We can thus hypothesise that to level the playing field in the most effective manner, we may need to provide both deductive instruction and mixed-modality input.

CONCLUSION

The present study investigated ATIs in light of current theorising of the construct of language learning aptitude. We compared the performance of international adult volunteers who were instructed in four different experimental conditions on two morphological features of Polish, a language they were entirely unfamiliar with. We developed four online language lessons that reflected real-world teaching and learning materials using a presentation–controlled practice approach. The treatment conditions differed in terms of input modality (auditory, written, and mixed) and the nature of the explicit instruction adopted (inductive and deductive).

The study resulted in several key findings. First, we replicated the potentially neutralising effect of deductive instruction that had been observed in previous research. We therefore conclude that such an approach may help level the playing field between learners with different cognitive abilities by compensating for weaknesses in language-analytic ability in particular.

Second, and arguably more importantly, we identified four distinct learner profiles: high aptitude, low aptitude, memory oriented, and analytically oriented. As the cluster analysis on which this finding is based has been, to the best of our knowledge, the first in the ATI research field for more than two decades, this is an intriguing and perhaps unexpected result. If the same clusters emerge in very different populations, in different learning contexts, and based on different measurement instruments, a strong case for generalisability can be made. Given that an understanding of learner types has not only theoretical but also practical implications, we are a step closer to potentially providing instructional conditions that can maximise L2 learning success. In today's context of online language learning tools, which allow for adaptive and individualised input, recommendations arising from ATI research are arguably much more actionable than they were even a decade ago. Therefore, the identification of learner profiles seems a timely and worthwhile endeavour with the potential to yield useful real-world implications.

Third, our findings suggest that input modality may be as important as the inductive-versus-deductive contrast in explicit instruction. In other words, it seems advisable to consider the role of learners' perceptual abilities and/or preferences (auditory and visual), over and above the role of cognitive abilities (memory and language analysis). This appears to be an underresearched area that future studies should aim to take into account.

It goes without saying that our study had its limitations, two of which are highlighted here. First, our participant sample was not as large as we would have wished, resulting in uneven numbers across experimental conditions and too few cases in some of the cells in the analysis of ATI effects. As a consequence, we cannot draw any meaningful conclusions about the possible role of a high-aptitude profile in the mixed inductive group, for example. Second, we did not include any delayed posttests due to the practical challenge of having to persuade participants to return yet again after completing lengthy sessions spread over several days.

Taken together with our main findings, these shortcomings point towards avenues for future research. As longer-term effects are of major interest in L2 learning research, a future study could aim for several testing points in the context of a time-series design. Moreover, it would ideally include an implicit or at least an incidental experimental condition in order to uncover any potential implicit aptitude effects, given that we currently know very little about the precise role implicit abilities might play. In view of the tantalising trends we observed with regard to the role of input modality, a future study could again manipulate this variable to confirm or reject the hypothesis that mixed-modality input is an important ingredient for neutralising the impact of individual learner differences. Last but not least, future studies should investigate different languages and target features. In sum, we believe the ATI agenda is wide open once more in light of recent theoretical and technological developments and is waiting to be addressed with renewed vigour.

ORCID

Karen Roehr-Brackin https://orcid.org/0000-0002-1152-5852
Renato Pavlekovic https://orcid.org/0000-0002-5955-0379

ENDNOTE

[1] A reviewer expressed concern about our use of the LLAMA suite. While the LLAMA has been subject to justified criticism (Bokander & Bylund, 2020), the new version 3 (https://llamatests.org) is much improved in terms of performance, reliability, and validity (see Rogers et al., 2023). Indeed, validation of the LLAMA suite is an ongoing endeavour. Data from 1,370 participants who have taken LLAMA version 3 are being analysed. Internal consistency (Cronbach's alpha) coefficients range

from 0.73 to 0.91 for individual subtests. Test–retest reliability has been assessed on the basis of a subsample of 94 participants who were tested and then retested after a 4-week interval. Findings show that the subtests correlated over time. The LLAMA suite as a whole yielded intraclass correlations of 0.64 (agreement) and 0.76 (consistency), which is deemed fair to good (Rogers et al., 2024). Moreover, from a purely practical perspective, the LLAMA suite is the only viable aptitude measure that is available to L2 researchers working with international multilingual samples (Roehr-Brackin, 2021), that is, samples that are representative of the majority of language learners worldwide.

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. https://doi.org/10.1017/S027226310808073X

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

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

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

Carroll, J. B., & Sapon, S. (1959). *The modern language aptitude test*. Psychological Corporation.

Cockcroft, K., Wigdorowitz, M., & Liversage, L. (2019). A multilingual advantage in the components of working memory. *Bilingualism: Language and Cognition*, *22*, 15–29. https://doi.org/10.1017/S1366728917000475

DeKeyser, R. M. (2005). What makes learning second-language grammar difficult? A review of issues. *Language Learning*, *55*, 1–25. https://doi.org/10.1111/j.0023-8333.2005.00294.x

DeKeyser, R. M. (2012). Interactions between individual differences, treatments, and structures in SLA. *Language Learning*, *62*, 189–200. https://doi.org/10.1111/j.1467-9922.2012.00712.x

DeKeyser, R. M. (Ed.). (2021). *Aptitude-treatment interaction in second language learning*. John Benjamins. https://doi.org/10.1075/bct.116

Doughty, C. J. (2019). Cognitive language aptitude. *Language Learning*, *69*, 101–126. https://doi.org/10.1111/lang.12322

Erlam, R. (2005). Language aptitude and its relationship to instructional effectiveness in second language acquisition. *Language Teaching Research*, *9*, 147–171. https://doi.org/10.1191/1362168805lr161oa

Grañena, G. (2013a). Cognitive aptitudes for second language learning and the LLAMA language aptitude test. In G. Grañena & M. H. Long (Eds.), *Sensitive periods, language aptitude, and ultimate L2 attainment* (pp. 105–129). John Benjamins.

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

Grañena, G. (2013c). Reexamining the robustness of aptitude in second language acquisition. In G. Grañena & M. H. Long (Eds.), *Sensitive periods, language aptitude, and ultimate L2 attainment* (pp. 179–204). John Benjamins.

Grañena, G. (2016). Cognitive aptitudes for implicit and explicit learning and information-processing styles: An individual differences study. *Applied Psycholinguistics*, *37*, 577–600. https://doi.org/10.1017/S0142716415000120

Grañena, G. (2020). *Implicit language aptitude*. Cambridge University Press. https://doi.org/10.1017/9781108625616

Grañena, G., & Yilmaz, Y. (2018). Aptitude-treatment interaction in L2 learning: A research synthesis. *Studies in English Education*, *23*, 803–830. https://doi.org/10.22275/SEE.23.4.02

Housen, A., Pierrard, M., & Van Daele, S. (2005). Structure complexity and the efficacy of explicit grammar instruction. In A. Housen & M. Pierrard (Eds.), *Investigations in instructed second language acquisition* (pp. 235–269). Mouton de Gruyter.

Housen, A., & Simoens, H. (2016). Introduction: Cognitive perspectives on difficulty and complexity in L2 acquisition. *Studies in Second Language Acquisition*, *38*, 163–175. https://doi.org/10.1017/S0272263116000176

Hulstijn, J. H., & de Graaff, R. (1994). Under what conditions does explicit knowledge of a second language facilitate the acquisition of implicit knowledge? A research proposal. *AILA Review*, *11*, 97–112.

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

Hwu, F., & Sun, S. (2012). The aptitude-treatment interaction effects on the learning of grammar rules. *System*, *40*, 505–521. https://doi.org/10.1016/j.system.2012.10.009

Juffs, A., & Harrington, M. (2011). Aspects of working memory in L2 learning. *Language Teaching*, *44*, 137–166. https://doi.org/10.1017/S0261444810000509

Kaufman, S. B., DeYoung, C. G., Gray, J. R., Jiménez, L., Brown, J., & Mackintosh, N. (2010). Implicit learning as an ability. *Cognition*, *116*, 321–340. https://doi.org/10.1016/j.cognition.2010.05.011

Kempe, V., & Brooks, P. J. (2011). Individual differences in adult second language learning: A cognitive perspective. *Scottish Languages Review*, *23*, 15–22.

Kidd, E., Donnelly, S., & Christiansen, M. H. (2018). Individual differences in language acquisition and processing. *Trends in Cognitive Sciences*, *22*, 154–169. https://doi.org/10.1016/j.tics.2017.11.006

Kim, K. M., & Godfroid, A. (2019). Should we listen or read? Modality effects in implicit and explicit knowledge. *Modern Language Journal*, *103*, 648–664. https://doi.org/10.1111/modl.12583

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

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

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*, 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*, 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. (2023). Working memory and second language learning: A critical and synthetic review. In A. Godfroid & H. Hopp (Eds.), *Routledge handbook of second language acquisition and psycholinguistics* (pp. 348–360). Routledge. https://doi.org/10.4324/9781003018872-32

Li, S., & DeKeyser, R. (2021). Implicit language aptitude: Conceptualizing the construct, validating the measures, and examining the evidence. *Studies in Second Language Acquisition*, *43*, 473–497. https://doi.org/10.1017/S0272263121000024

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

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

Linck, J. A., & Weiss, D. J. (2011). Working memory predicts the acquisition of explicit L2 knowledge. In C. Sanz & R. P. Leow (Eds.), *Implicit and explicit language learning: Conditions, processes, and knowledge in SLA and bilingualism* (pp. 101–113). Georgetown University Press.

McLaughlin, B. (1995). Aptitude from an information-processing perspective. *Language Testing*, *12*, 370–387.

Meara, P. (2005). *LLAMA language aptitude tests*. Lognostics.

Meara, P., & Rogers, V. (2019). *The LLAMA tests v.3*. Lognostics.

Miyake, A., & Friedman, N. P. (1998). Individual differences in second language proficiency: Working memory as language aptitude. In A. F. Healy & L. E. Bourne (Eds.), *Foreign language learning: Psycholinguistic studies on training and retention* (pp. 339–364). Erlbaum.

Norris, J. M., & Ortega, L. (2001). Does type of instruction make a difference? Substantive findings from a meta-analytic review. *Language Learning*, *51*, 157–213. https://doi.org/10.1111/j.1467-1770.2001.tb00017.x

Pavlekovic, R., & Roehr-Brackin, K. (2024). Aptitude for explicit and implicit learning. *Journal of the European Second Language Association*, *8*, 1–17. https://doi.org/10.22599/jesla.117

Peirce, J., Gray, J. R., Simpson, S., MacAskill, M., Höchenberger, R., Sogo, H., Kastman, E., & Lindeløv, J. K. (2019). PsychoPy2: Experiments in behavior made easy. *Behavior Research Methods*, *51*, 195–203. https://doi.org/10.3758/s13428-018-01193-y

Perruchet, P. (2021). Why is the componential construct of implicit language aptitude so difficult to capture? A commentary on the special issue. *Studies in Second Language Acquisition*, *43*, 677–691. https://doi.org/10.1017/S027226312100019X

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). John Benjamins.

Ranta, L. (2008). Aptitude and good language learners. In C. Griffiths (Ed.), *Lessons from good language learners* (pp. 142–155). Cambridge University Press.

Robinson, P. (2001). Individual differences, cognitive abilities, aptitude complexes and learning conditions in second language acquisition. *Second Language Research*, *17*, 368–392. https://doi.org/10.1177/026765830101700405

Robinson, P. (2005). Aptitude and second language acquisition. *Annual Review of Applied Linguistics*, *25*, 46–73. https://doi.org/10.1017/S0267190505000036

Robinson, P. (2007). Aptitudes, abilities, contexts, and practice. In R. M. DeKeyser (Ed.), *Practice in a second language: Perspectives from applied linguistics and cognitive psychology* (pp. 256–286). Cambridge University Press.

Roehr-Brackin, K. (2021). Measuring aptitude. In P. Winke & T. Brunfaut (Eds.), *The Routledge handbook of second language acquisition and language testing* (pp. 147–156). Routledge.

Roehr-Brackin, K., Loaiza, V., & Pavlekovic, R. (2023). Language learning aptitude in older adults. *Journal of the European Second Language Association*, *7*, 1–15. https://doi.org/10.22599/jesla.93

Rogers, V., Bokander, L., Galvin, T., Ray, R., Rogers, B., & Meara, P. (2024). Developments in LLAMA test validation. Paper presented at the 33rd Conference of the European Second Language Association. Montpellier, France.

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

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.

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

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

Serafini, E. J. (2017). Exploring the dynamic long-term interaction between cognitive and psychosocial resources in adult second language development at varying proficiency. *Modern Language Journal*, *101*, 369–390. https://doi.org/10.1111/modl.12400

Serafini, E. J., & Sanz, C. (2016). Evidence for the decreasing impact of cognitive ability on second language development as proficiency increases. *Studies in Second Language Acquisition*, *38*, 607–646. https://doi.org/10.1017/S0272263115000327

Skehan, P. (1986). Cluster analysis and the identification of learner types. In V. Cook (Ed.), *Experimental approaches to second language learning* (pp. 81–94). Pergamon.

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

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

Skehan, P. (2015). Foreign language aptitude and its relationship with grammar: A critical overview. *Applied Linguistics*, *36*, 367–384. https://doi.org/10.1093/applin/amu072

Skehan, P. (2016). Foreign language aptitude, acquisitional sequences, and psycholinguistic processes. In G. Grañena, D. O. Jackson, & Y. Yilmaz (Eds.), *Cognitive individual differences in second language processing and acquisition* (pp. 17–40). John Benjamins.

Spada, N., & Tomita, Y. (2010). Interactions between type of instruction and type of language feature: A meta-analysis. *Language Learning*, *60*, 263–308. https://doi.org/10.1111/j.1467-9922.2010.00562.x

Staples, S., & Biber, D. (2015). Cluster analysis. In L. Plonsky (Ed.), *Advancing quantitative methods in second language research* (pp. 243–274). Routledge.

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

Unsworth, N., Heitz, R. P., Schrock, J. C., & Engle, R. W. (2005). An automated version of the operation span task. *Behavior Research Methods*, *37*, 498–505. https://doi.org/10.3758/BF03192720

Vatz, K., Tare, M., Jackson, S. R., & Doughty, C. J. (2013). Aptitude-treatment interaction studies in second language acquisition: Findings and methodology. In G. Grañena & M. Long (Eds.), *Sensitive periods, language aptitude, and ultimate attainment* (pp. 273–292). John Benjamins.

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

Wesche, M. B. (1981). Language aptitude measures in streaming, matching students with methods, and diagnosis of learning problems. In K. C. Diller (Ed.), *Individual differences and universals in language learning aptitude* (pp. 119–154). Newbury House.

Williams, J. N. (2012). Working memory and SLA. In S. M. Gass & A. Mackey (Eds.), *The Routledge handbook of second language acquisition* (pp. 427–441). Routledge.

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

SUPPORTING INFORMATION

Additional supporting information can be found online in the Supporting Information section at the end of this article.

How to cite this article: Roehr-Brackin, K., Baranowska, K., Pavlekovic, R., & Scheffler, P. (2024). The role of individual learner differences in explicit language instruction. *Modern Language Journal*, *108*, 815–845. https://doi.org/10.1111/modl.12963