System 131 (2025) 103656

Contents lists available at ScienceDirect

System

journal homepage: www.elsevier.com/locate/system

Implicit statistical learning and working memory predict EFL development and written task outcomes in adolescents

Diana Pili-Moss [a,*], Phillip Hamrick [b], Katharina Wendebourg [c], Torben Schmidt [a], Detmar Meurers [d]

[a] *Institute of English Studies, Leuphana University Lüneburg, Germany*
[b] *Department of Psychological Science, Kent State University, United States*
[c] *Hector Institut für Empirische Bildungsforschung, University of Tübingen, Germany*
[d] *Leibniz Institut für Wissensmedien, University of Tübingen, Germany*

A R T I C L E I N F O

Keywords:
Explicit language aptitude
Implicit language aptitude
Working memory
EFL grammar
Written L2 task
CALL
Explicit long-term memory
Statistical learning ability

A B S T R A C T

Investigating the relationship between cognitive individual differences and second language learning has been central to second language acquisition research conducted in controlled laboratory conditions and in educational instructed contexts. However, not much research to date has simultaneously explored the role of multiple cognitive abilities for L2 development or task outcomes in educational environments. In the present study, 77 secondary-school EFL learners engaged in intensive digital practice of direct questions, alongside regular classroom instruction, for a period of two and a half weeks. They completed digital pretests and, at the end of the instruction and practice period, were administered digital posttests and a pen-and-paper communicative written task. Measures of the learners' declarative memory, implicit statistical learning and working memory capacity were taken. Mixed-effect and multiple regression models revealed that a positive interaction between implicit statistical learning and working memory capacity predicted both posttest scores and task outcomes, whereas declarative memory did not significantly relate to either measure. It is suggested that the synergetic relationship between implicit statistical learning and working memory capacity may be key to the process of updating morphosyntactic representations of direct questions with positive effects for both L2 development and use.

1. Introduction

The study of cognitive individual differences (IDs) as components of second language (L2) aptitude has benefitted from advances in two main research strands, language aptitude studies investigating L2 learning in educational instructed contexts (e.g., Li, 2016), and laboratory-based training studies, that have mostly drawn their methodology from cognitive psychology (e.g., Hamrick et al., 2018).

In recent language aptitude research, explicit language aptitude (explicit aptitude) is a tripartite construct (Skehan, 2002) including explicit associative long-term memory (rote learning), phonetic coding ability and language analytic ability, assessed using aptitude testing batteries such as MLAT (Carroll & Sapon, 2002), LLAMA (Meara & Rogers, 2019) and HiLAB (Linck et al., 2013). By

This article is part of a special issue entitled: Individual Differences and TBLT published in System.

* Corresponding author.
E-mail address: diana.pili-moss@leuphana.de (D. Pili-Moss).

https://doi.org/10.1016/j.system.2025.103656
Received 30 May 2024; Received in revised form 22 February 2025; Accepted 9 March 2025
Available online 15 March 2025

contrast, laboratory-based studies have mainly operationalized explicit aptitude as declarative memory, a type of long-term memory that supports fast and context-independent learning of facts, episodic and semantic information and explicit associations (Ullman, 2020). Although the two operationalizations are not fully equivalent, they largely overlap with respect to the associative memory component.

Implicit language aptitude (implicit aptitude) is a comparatively more recent construct that broadly denotes the ability to acquire an L2 implicitly, by virtue of iterated exposure to relevant language input. Similar to explicit aptitude, it is componential in nature. Two main components are procedural learning ability (the ability to implicitly acquire linguistic rule-based patterns such as word order or inflectional morphology and to efficiently proceduralize language in use; e.g., N. Ellis, 2015; Ullman, 2020) and statistical learning ability (the ability to track the linguistic input and extract statistical regularities from its distributional properties; e.g., Misyak & Christiansen, 2012). Although a subtest of the LLAMA test (LLAMA-D) has sometimes been used as an index of implicit aptitude, studies have generally adopted paradigms developed in the field of cognitive psychology to measure it, such as serial reaction time (SRT), implicit statistical and probabilistic learning tasks. Following recent SLA approaches to the definition of L2 aptitude constructs (e.g., Li & DeKeyser, 2021; Morgan-Short et al., 2014), we will include aptitude measures employed in laboratory-based cognitive research as indexes of components of explicit and implicit aptitude. Specifically, we will use a measure of declarative memory as an index of explicit aptitude and a measure of implicit statistical learning as an index of implicit aptitude.

Although declarative memory does not necessarily imply awareness (Ullman, 2020), both declarative memory and explicit associative memory are largely overlapping cognitive functions assessed via similar (or sometimes the same) tests. In what follows we will employ one or the other term interchangeably, depending on how they were used in reviewed individual studies.

Beside measures of explicit and implicit aptitude, working memory has been identified as a key cognitive predictor of L2 learning and development (Williams, 2015; for a recent review see e.g., Schwieter & Wen, 2022). Working memory (WM) is a limited capacity cognitive function enabling short-term encoding, processing and manipulation of sensory information. SLA research has mainly conceptualized WM as a modular system including sensory-specific short-term storage components, a central executive component and an episodic buffer, providing an interface with the long-term memory system (Baddeley, 2012). Alternative conceptualizations have defined WM as a temporary activation of long-term memory representations through focal attention (e.g., Cowan, 2008).

Independently of the specific WM model adopted, it has been suggested that a close 'two-way' relationship between working memory and long-term memory exists, with stronger WM capabilities supporting the efficiency of long-term memory processing (Cotton and Ricker, 2021; Park et al., 2020; but see e.g., Pedraza et al., 2024 for different findings). Compatibly with this hypothesis, stronger WM capabilities would implicate more efficient transfer of (new) L2 information from WM to long-term memory. Conversely, enhanced WM capabilities would also support more efficient online processing of information retrieved from long-term memory in language use (e.g., in L2 production).

Importantly, although a number of studies have investigated working memory as a main predictor of L2 development alongside memory-based measures of explicit and implicit aptitude (e.g., Faretta-Stutenberg & Morgan-Short, 2018; Linck et al., 2013; McDonough & Trofimovich, 2016; Morgan-Short et al., 2014), potential moderation (interaction) effects between cognitive predictors have generally not been considered. The present study investigates the extent to which declarative memory (an index of explicit aptitude), implicit statistical learning (an index of implicit aptitude) and working memory capacity predict EFL development and use in a school-based environment. However, beside examining the relationship of individual factors to the outcome measures, the study also explores potential moderation effects, i.e., the extent to which a relationship between a given predictor and L2 outcomes significantly increases or decreases the strength of the relationship of a second predictor with the same outcomes.

1.1. Cognitive IDs in laboratory-based training studies and educational environments

SLA research conducted in the laboratory and in educational environments has shown that cognitive individual differences are key predictors of L2 outcomes (Hamrick et al., 2018; Kersten & Winsler, 2023; Li, 2016). For example, laboratory-based training studies have evidenced strong relationships between L2 learning of morphosyntax and declarative memory in the early stages of learning, with the relationship becoming less robust at higher stages of L2 proficiency or exposure (e.g., Hamrick, 2015; Morgan-Short et al., 2014; for the relationship between declarative memory and L2 vocabulary learning see e.g., Murphy et al., 2021). Beside amount of exposure and proficiency level, other modulating factors that have been identified are age (children rely on declarative memory to a lesser extent compared to adults; Pili-Moss, 2021), type of instruction (declarative memory is more strongly associated with explicit rather than incidental instruction; e.g., Brill-Schuetz & Morgan-Short, 2014), or type of outcome measure (measures of sentence comprehension remain positively related to declarative memory even at higher proficiency levels; e.g., Pili-Moss et al., 2020).

For educational contexts, Li (2016) reported a significant, though weak, relationship between rote memory and overall L2 learning (r = 0.24) in a meta-analysis of 66 aptitude studies with adolescent and adult participants. Significant positive associations also emerged for L2 grammar learning, $r = 0.19$, and literacy skills except writing, $0.19 < r < 0.21$. Some studies not included in the meta-analysis reported positive associations between explicit associative memory and outcomes for L2 listening and reading in highly-proficient adults (e.g., Linck et al., 2013), whereas others did not find that declarative memory predicted L2 grammar outcomes in adults instructed in the classroom (e.g., Faretta-Stutenberg & Morgan-Short, 2018). However, the predictive role of declarative memory for the development of adult L2 grammar and vocabulary was confirmed by studies where L2 instruction was provided through custom-designed or commercial digital platforms (e.g., Duolingo; Vasileiou & Pili-Moss, 2022). Overall, investigations in educational environments have obtained results that are, with some exceptions, largely compatible with the findings of laboratory-based training studies, pointing to a substantial role of explicit associative or declarative memory in adult instructed L2. Compared to the adolescent and adult literature, the few child studies conducted to date return less clear-cut results, with some studies finding positive relationships between learning of sound-symbol associations and learning of L2 vocabulary and grammar (e.g., Tellier & Roehr-Brackin, 2013) and others where these relationships were not confirmed (e.g., Roehr-Brackin & Tellier, 2019).

With regard to implicit aptitude, laboratory training studies have found that one type of implicit aptitude, procedural learning ability (measured by SRT or probabilistic tasks), is a significant positive predictor of language automatization in adults (Pili-Moss et al., 2020) and generally tends to be more strongly associated with language accuracy at late stages of learning (e.g., Morgan-Short et al., 2014). By contrast, L2 learning in children has been showed to rely on procedural learning ability from the early stages of exposure to a new language (Pili-Moss, 2021). In laboratory studies, statistical learning ability has also been found to positively relate to L2 grammar outcomes (e.g., Brooks et al., 2017; McDonough & Trofimovich, 2016; Misyak & Christiansen, 2012).

In educational environments, procedural learning ability was found to be a positive predictor of L2 exam scores in adolescents (Kaufmann et al., 2010), of L2 grammar gains in adult immersion contexts (e.g., Bolibaugh & Foster, 2021; Faretta-Stutenberg & Morgan-Short, 2018), of adult L2 speech performance (e.g., Granena, 2019), and of adult L2 literacy skills at high proficiency levels (e.g., Linck et al., 2013). Statistical learning ability was also associated to better L2 outcomes in child receptive vocabulary (e.g., Ren & Wang, 2023) and adolescent reading skills (Frost et al., 2013).

Studies investigating the predictive role of WM capacity for L2 learning have been characterised by mixed findings across different bodies of research, which makes identifying robust trends complex (for reviews of studies adopting WM capacity or storage measures, see e.g., Li, 2022). Although some experimental training studies found that working memory capacity predicts L2 grammar learning (at least in explicit instruction conditions, e.g., Tagarelli et al., 2011), others did not find evidence for this association on measures of WM storage or capacity (e.g., Brooks et al., 2017; McDonough & Trofimovich, 2016; Morgan-Short et al., 2014; Pili-Moss, 2021, among others). For nonexperimental SLA studies, the meta-analysis in Linck et al. (2014) revealed overall positive but weak associations between WM and measures of L2 comprehension and production ($r = 0.25$). This result is comparable to the one reported in Li (2017), a meta-analysis of L2 interactional research ($r = 0.23$).

A further aspect that deserves mention concerns the investigation of the combined effects of multiple types of cognitive individual differences. Though a number of studies have investigated the individual effects of sets of cognitive predictors (or covariates), very few have tested the effects of cognitive predictors' interactions for L2 outcomes. Based on the hypothesis that, beyond individual effects, cognitive abilities are likely to show patterns of interaction akin to those observed between memory functions in neurophysiological research (e.g., Poldrack & Packard, 2003), some studies in this area have started to explore moderation effects between cognitive ability measures. Initial evidence shows that declarative and procedural memory positively (co-operatively) interact supporting adult L2 automatization (Pili-Moss et al., 2020) and grammar learning (Pili-Moss, 2022) and that declarative memory positively interacts with the amount of previous knowledge in L2 incidental vocabulary learning (Murphy et al., 2021).

Overall, extant research in instructed L2 learning indicates an important role of the declarative/explicit associative component of explicit aptitude. Implicit aptitude positively predicts L2 outcomes in instructed environments as a function of L2 proficiency, sustained exposure to L2 input and immersion experience. There is also initial experimental evidence that the relationship between language aptitude and L2 outcomes is moderated by age, although more research is needed. Evidence regarding the association of WM ability with L2 development is mixed and requires further investigation. Finally, although the L2 instruction administered in laboratory training studies has been mainly computer-mediated, very few studies to date have gauged the effects of cognitive ability in educational digital environments, a gap in research that needs addressing in view of the fast-spreading adoption of digital L2 instruction in mainstream education (Schmidt & Strasser, 2022).

1.2. *Cognitive individual differences in task-based L2 instruction*

Among the studies that have elucidated the relationship between cognitive individual differences and L2 outcomes in educational instructed contexts, some have explored how cognitive ability contributes to L2 development, attainment and performance in task-based instruction. The cognitive ability that has been most thoroughly studied in this context is WM, indexed by measures of its central executive or short-term storage components. Theoretical models of cognitive engagement in tasks (e.g., Kormos, 2012; Skehan, 2002) have generally predicted a positive relationship between WM abilities and L2 task outcomes, although research findings have not consistently supported this association.

In a recent narrative synthesis of research investigating the role of working memory in L2 oral interaction, An and Li (2022) reported mixed supporting evidence for a positive relationship between WM and learners' behaviour in tasks (e.g., noticing of oral corrective feedback), weak associations between WM and oral fluency and mixed results for the association of WM and the efficacy of oral corrective feedback. In written tasks, WM ability was found to significantly positively relate to measures of learners' behaviour (e.g., Révész et al., 2023), as well as to measures of complexity (e.g., Mavrou, 2020; Yi & Ni, 2015; but see Vasylets & Marín, 2021; Zabildea, 2017, for different results), accuracy (e.g., Mavrou, 2020; Vasylets & Marín, 2021; see Zahibi, 2018; for different findings), and fluency (e.g., Kim et al., 2021; Zahibi, 2018). In at least one study (Manchón et al., 2023) WM was not significantly related to any of the CAF measures.

Even if to a lesser extent compared to WM, aspects of explicit aptitude have been investigated in L2 tasks. Studies that employed oral tasks mainly explored explicit aptitude relative to the efficacy of different types of corrective feedback techniques. For example, Li (2015) found that language analytic ability (measured by the Words in Sentences section of the MLAT) predicted morphological learning gains in learners who had received metalinguistic feedback during oral narration and interview tasks. Studies that employed written tasks have found associations between task outcomes and explicit aptitude indexed by measures of grammatical sensitivity or aptitude battery scores. For example, Kormos and Trebits (2012) found that grammatical sensitivity was positively related to clause length in a written description task. Similarly, metalinguistic and morphological awareness were respectively linked to teachers' rating of written production in Kormos and Sáfár (2008) and to a range of writing quality indexes, including accuracy, in Peng et al. (2021).

Vasylets et al. (2022) analysed the relationship between LLAMA aptitude scores and CAF outcomes in a written problem-solving task, finding positive statistically significant relationships only for grammar inferencing ability. Significant relationships between subtests of the LLAMA battery gauging aspects of explicit associative memory and L2 writing outcomes were also found in Mujtaba et al. (2021) for LLAMA-B (vocabulary learning), as well as in Mujtaba et al. (2021) and Yang et al. (2019) for LLAMA-E (sound-symbol correspondence).

In one of the rare studies to investigate the relationship between implicit language aptitude and task-based learning to date, Granena (2019) used semantic priming to explore the relationship between implicit learning ability and CAF measures in a speaking task. The study found that priming was positively related to L2 speed fluency, a finding that the author related to the hypothesized learners' enhanced ability to retrieve implicit long-term memory knowledge during on-line speech performance.

To summarise, a number of studies have started to elucidate the relationship between cognitive ability and task performance. However, most studies considered exclusively WM measures and only a handful have explored explicit aptitude. Almost no research is available on the role of implicit aptitude in tasks and no studies to date have specifically explored the relationships between written task performance and implicit aptitude.

1.3. *The present study*

The aim of the present study was to investigate the relationship between a set of language aptitude measures (explicit aptitude, measured by declarative memory in visual sequence learning; implicit aptitude, measured by implicit statistical learning ability in visual sequence learning; and WM capacity, measured by a backward digit span task) and the development and use of EFL direct questions in L1 German secondary school learners (12-13-year olds). After two and a half weeks of instruction, including regular classroom hours and additional digital practice, EFL outcomes were measured by: (a) pre-posttest accuracy gains in a digitally administered fill-in-the-gap test and (b) proficiency in free question production in a pen-and-paper written communicative task. The research questions were formulated as follows.

RQ1: To what extent do declarative memory, implicit statistical learning ability and WM capacity predict morphosyntactic development in L2 direct questions as measured by posttest scores?
RQ2: To what extent do declarative memory, implicit statistical learning ability and WM capacity predict morphosyntactic proficiency in the free production of L2 direct questions in the written task?
RQ3: To what extent do interactions between the cognitive variables of interest predict (a) posttest scores, and (b) written task outcomes?

Considering that the learners are adolescents in an instructed environment, we hypothesize that declarative memory may be a significant predictor of both posttest and task outcomes. However, since direct questions are grammar constructions these learners will have encountered multiple times in the course of previous school years, and that they are intensively exposed to multiple exemplars of the construction in the digital practice in the present treatment, a role of implicit aptitude in supporting learning cannot be excluded. WM may likewise support EFL question development and use. This could be particularly the case in the written task, where question production occurs under conditions of communicative pressure (see Methods). Concerning RQ3, previous studies found that positive interactions between declarative and procedural memory measures significantly predicted L2 development (e.g., Pili-Moss, 2022), hence similar effects between declarative memory and statistical learning ability are investigated in the present study. Further, under the hypothesis that WM supports processing of long-term stored linguistic information, positive interactions between WM and long-term memory measures would also be expected.

2. Methods

2.1. *Participants*

The participants were 77 L1 German 12-13-year olds, from three Grade 7 intact classes (27, 24, and 27 participants, respectively) at one academic-track secondary school in Northern Germany, who had started to learn EFL in the classroom from primary school (Grade 3). Although each class group was taught by a different English teacher, the teachers planned the yearly syllabus as a team and delivered teaching and assessments in parallel in the three classes, thus ensuring group comparability in terms of the classroom instruction conditions. Random effects of class group were included in all statistical analyses to make sure additional random differences due to grouping were controlled for.

The study was approved by the Ethics Office of the University of Tübingen and the Ministries of Education of Baden-Württemberg and Lower Saxony. Parental consent to participate was obtained for 75 participants prior to the beginning of the study. The two students for which consent was not available participated in all instructional activities and their data was excluded from analysis.

2.2. *Study design and procedure*

The present study was part of a larger school-based investigation conducted over a whole school year (Fig. 1). Here we report on a subset of data collected during three weeks between February and March 2022 (Cycle 3).

Alongside classroom instruction, learners completed exercises on the FeedBook, a web-based intelligent tutoring system offering feedback-supported EFL digital practice (Meurers et al., 2019). Classroom instruction and digital practice opportunities were the same across the three class groups. Overall, during this period learners received 8 h of regular classroom-based instruction, 3 h the first two weeks and 2 h the third week. During the last 2 h learners engaged in a written communicative target-task (for open access to the complete lesson plans and class materials see Pili-Moss et al., 2023). Additionally, learners could practice on the platform for about 45 min four times a week during the first two weeks. Learners were administered digital pre- and posttests immediately prior to the beginning and after the end of the Cycle 3 instruction and practice period. Due to the school's organizational requirements, the cognitive tests were administered in school in a quiet room in supervised laboratory conditions during a two-day session in June 2022. They were programmed by the first author on Gorilla and administered individually in a single paradigm on a DELL personal computer (about half an hour in total). The order of tasks (implicit statistical learning, working memory capacity, declarative memory) was the same for all participants.

2.3. Classroom instruction

Classroom instruction was custom-designed collaboratively by one of the teachers with members of the research team. Cycle 3 (grammar focus on present and past simple direct questions and the topic of friends and feelings) followed a backward planning model where classroom instruction, pre-communicative activities and form-focused digital practice served the purpose of building up language knowledge and skills supporting learners in the final written communicative task. Given the role of form-focused activities and explicit language learning, the overall instructional approach reflects a task-supported learning model rather than a task-based language teaching approach (Ellis, 2021).

The written communicative task (Appendix A1-2) foresaw different phases where learners engaged in individual, pair, and plenary work. To begin with, learners were paired and each pair presented with either Version (a) or (b) of the task. The worksheet presented only a visual cue (a picture of a teenager) and a title (respectively '*Bad decision!*' or '*It's definitely not my fault!*'). The initial activity instructed learners to consider the cues and individually write down a list of exploratory questions with the aim to gain more information about the protagonists and what had happened to them. Although the initial activity was not explicitly timed, transitions between task stages were teacher-led and the question list quickly fed into a second phase in the communicative activity. Once the questions were drafted, a pair completing the same version of the task swapped worksheets with a pair completing the alternate version and learners in each pair collaborated in writing down possible answers to the questions. Worksheets were swapped again, and, after a brainstorming plenary phase, each student went on to individually write down their own version of the story based on the answers they had received and their own ideas. Finally, recordings were played where learners could listen to what had really happened from the

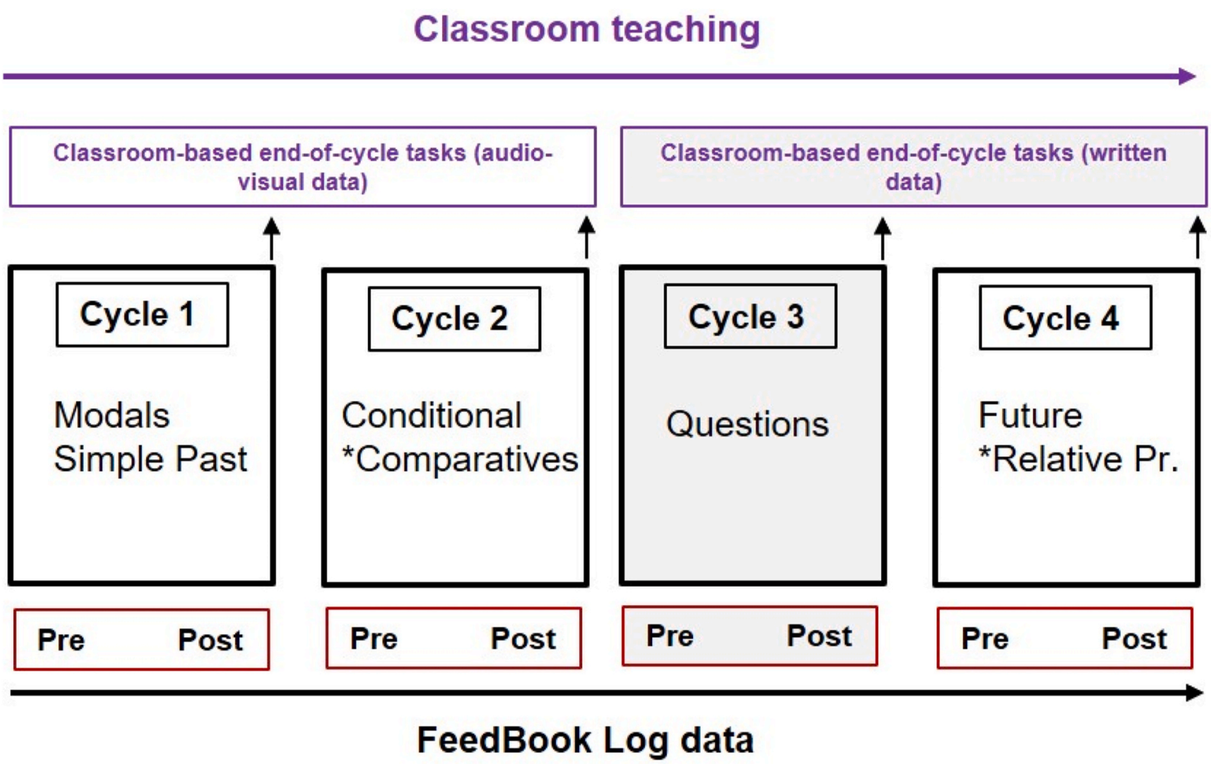

Fig. 1. Research design. Cycle 3 (in grey) represents the focus of the present study.

voices of the two protagonists.

2.4. Digital practice

Digital practice consisted of 56 exercises (fill-in the blank, multiple choice, and open answers) that were created collaboratively by a teacher and members of the research team and accessible via school-provided tablets. During digital practice, learners were free to complete the exercises at their own pace, resulting in differences in the number of exercises attempted across learners. In response to incorrect answers, the platform offered form-focused feedback in English (metalinguistic cues) or presented the response 'This is not what I am expecting. Please try again' if the response did not match an error category recognised by the system. Feedback on correct answers was provided in the form of a green tick.

2.5. Outcome measures

2.5.1. Pre- and posttest scores

Pretest (Cronbach's alpha = 0.868) and posttest (Cronbach's alpha = 0.882) contained 30 fill-in-the-blank items each and were constructed by the first author as equivalent instruments (15 questions in the present simple, 15 questions in the past simple, 15 wh- and 15 yes/no questions; Appendix B1-2). Learners were instructed that solutions could require up to three words, but not told how many at item level. Scores corresponded to the number of accurate item responses (maximum 30 points).

2.5.2. Target-task score (TT3 score)

For the purposes of defining an outcome measure, written data from the first phase of the target task (the question list) were analysed. A composite score was obtained by standardizing and averaging four attainment components (see Appendix A.4 for descriptions and a rationale), of which three were positive (overall number of questions attempted; correct number of question in the simple past; correct number of questions with auxiliaries do/does/did), and one was negative (number of grammatically inaccurate questions). After 22 % of the data was coded by the first author and a research assistant (interrater reliability 92.5 %), the first author completed the coding.

2.6. Cognitive measures

2.6.1. Implicit statistical learning ability

Implicit statistical learning ability was measured by means of a visual implicit sequence learning task (Cronbach's alpha = 0.446) adapted from Siegelman, Bogaerts, and Frost (2017). All task instructions appeared on-screen in German. Participants were instructed to pay attention to a series of complex shapes appearing one after the other on-screen for 1000 ms, separated by a fixation cross (250 ms), and informed that they would be asked some questions about the individual shapes later on. No reference to the shape sequencing was made, nor to the existence of a specific order. Unbeknownst to the participants, the continuous stream of 24 different shapes consisted of 8 fixed sequences of three shapes (triplets), half characterised by high transitional probability (TP = 1) and half characterised by lower transitional probability (TP = .33).

Overall, participants were exposed to 24 presentations of the 8 triplets in a pseudo-randomised order (192 presentations in total). To ensure they remained on task, the presentation was split in 12 blocks separated by attention checks requiring to react as quickly as possible by pressing the space bar. Not responding to any two attention checks within 10 s would result in exclusion from the experiment (no exclusions based on this criterion). Half-way through the task, participants were also invited to take a break of a few seconds if they felt tired. Immediately after exposure, participants were asked whether they had noticed any patterns in the order in which the shapes appeared, but none of them reported noticing any specific one.

Participants were then informed that the order in which shapes were presented was not random and sequence knowledge was tested presenting full triplets and pairs of shapes in the same modality employed in the initial exposure (i.e., one individual shape after the other). After each presentation of a triplet or pair, participants indicated whether they thought it represented a pattern sequence by clicking yes/no options on-screen. There were 32 test trials (16 sequences, half triplets and half pairs, and the corresponding 16 foils, presented in a pseudo-randomised order). Implicit statistical learning scores corresponded to the sum of correct trials (maximum score 32 points). Participants performed significantly above chance at group level (t [72] = 6.68, $p < .001$; chance = 50 %); 55 participants performed above chance, 7 participants at chance, and 11 below chance.

2.6.2. Declarative learning ability

Declarative memory was measured by means of a visual explicit sequence learning task (Kenanidis et al., 2024). In this task (Cronbach's alpha = 0.781), originally conceived as an explicit counterpart to Siegelman, Bogaerts, and Frost (2017), participants were exposed to a new set of 24 complex shapes organised in 8 triplets. Each triplet sequence was presented for 1000ms separated from the next by a fixation cross (250ms). Unlike the implicit version, participants were explicitly told to memorise the sequences and that they would be tested on how well they could remember them. Overall, the presentation consisted of 94 trials in 4 blocks of 16 pseudo-randomised trials in which each triplet was repeated twice. Blocks were separated by attention-check trials as described previously. Test trials were administered similarly to the previous task and included 32 items presented in a pseudo-randomised order (8 sequence triplets, 8 sequence pairs, and 16 corresponding foils). Declarative learning ability scores corresponded to the sum of correct trials (maximum score 32 points). Participants performed significantly above chance at group level (t [70] = 6.52, $p < .001$;

chance = 50 %); 53 participants performed above chance, 5 at chance, and 13 below chance.

2.6.3. WM

WM capacity was measured by a Backward Digit Span task (Massonnié et al., 2019). In this task participants are presented on-screen with a series of digit sequences of increasing length and asked to type the digits in the reversed order of display by clicking on an on-screen numeric pad. After two practice trials, each sequence was presented on-screen for 2 s following a fixation cross (750ms). The digit span ranged from 2 to 7 digits, and there were two trials for each sequence length. The task ended when participants made mistakes on both trials testing a given span length or for correct reproduction of at least one 7-digit sequence. The final score corresponded to the maximum span achieved.

2.7. Covariates

2.7.1. EFL proficiency

For the three class groups investigated, the school made available the overall EFL marks recorded at the end of the previous school year (Grade 6; CEFR target level A2). In the German system, marks range from 1 (highest) to 6 (lowest), so that numerically lower marks correspond to better attainment (for an interpretation of the scores see Appendix A.3). A one-way ANOVA ($F(2, 73) = 5.644$, $p = .005$), revealed a significant difference emerging between two of the three groups. In order to control for these differences, proficiency was included as a covariate in all statistical analyses.

2.7.2. Digital practice variables

Time. Since amount of digital practice varied across participants and constituted a confounding factor in the evaluation of the effects of the cognitive predictors, we used individual time spent on digital practice as a covariate (see also Hui et al., 2023) To obtain estimates, we used exercise time stamps provided by the system, calculated the average time (minutes) spent practising per day and summed these values.

2.7.3. Gains

Since learning during the instruction period varied across participants, pre-posttest gains were added as a covariate in the RQ2 – RQ3b analysis in order to avoid confounds in evaluating the effects of the cognitive predictors.

3. Results

3.1. Data analysis

The inferential analysis employed (binomial) linear mixed-effect models (*glmer* and *lmer* functions), and simple regression models (*lm* function) from the *lme4* package in R; Bates et al., 2015; R Core Team, 2021. In the models' derivations all continuous variables were standardized. Fixed effects (main effects and interactions) were added one at a time successively comparing (nested) models using the likelihood ratio test and the Akaike Information Criterion (AIC) or Adjusted R^2 values. Only interactions between the main cognitive predictors of interest were investigated. Fixed effects were included in a model if it converged and fit the data statistically significantly better compared to an equal (nested) model that did not include that variable, or at least had a comparably lower AIC. Statistically significant interactions were always included. For mixed-effects models, once the structure of fixed effects had been determined, random effects were explored in the same way starting from random effects on intercepts (test items, participants and class group in the RQ1-RQ3a analysis, and participants and class group in the RQ2-RQ3b analysis) and subsequently considering random effects on the slopes of the fixed effects.

In the RQ1-RQ3a analysis the outcome measure was accuracy in the Cycle 3 posttest and the independent variables included the three cognitive predictors of interest, pretest scores and the covariates listed in Methods, except Gains. In the RQ2-RQ3b analysis the outcome measure was use of questions in the target task measured by the calculated composite index, and the same set of independent

Table 1
Pre-posttest scores, target task scores and main predictors and covariates.

	M (*SD*)	*C.I (95 %)*	*SE*	*N*
Pretest	2.3 (3.7)	[2.2, 2.4]	0.62	71
Posttest	6.6 (5.6)	[6.4, 6.8]	0.10	58
Task (TT3)	0.0 (0.6)	[-0.0, 0.0]	0.01	65
DECL	20.1 (5.6)	[19.9, 20.3]	0.09	71
SL	18.8 (3.6)	[18.7, 19.0]	0.06	73
WM	4.4 (1.6)	[4.3, 4.4]	0.03	72
EFL$^+$	3.0 (0.8)	[3.0, 3.0]	0.01	75
Time (min)	69.2 (67.5)	[67.0, 71.5]	1.15	74
Gains	4.3 (4.9)	[4.2, 4.5]	0.08	56

Note. $^+$Higher positive values in the scale correspond to lower marks. DECL = declarative memory; SL = implicit statistical learning; WM = working memory.

variables, including Gains.

The interpretation of the effect sizes (R^2) follows the field-specific recommendations in Plonsky and Ghanbar (2018). Computation of the condition numbers led to exclude multicollinearity issues and inspection of the QQ-plots indicated that the distribution of residuals was compatible with a normal distribution. The R code relative to the models and the dataset are available open-access [https://osf.io/k7qwv/?view_only=d049c85117f14787bb66820447ec1881].

3.2. Descriptive statistics

Table 1 reports the descriptive statistics relative to pre-and posttest scores, target task scores, as well as the cognitive predictors and the covariates. A summary of the pairwise Pearson's correlations for all independent variables is in Appendix C.1.

3.3. RQ1-RQ3a

The first analysis explored the relationship between cognitive predictors and language outcomes in the digital posttest. The best-fitting model (Table 2a) accounted for about 36 % of the variance and included positive significant effects of Pretest ($OR = 2.01$, a small to medium effect), Time ($OR = 1.61$, a small effect), and an Implicit Statistical Learning × WM interaction ($OR = 1.42$, a small effect; SL and WM conditional effects $OR = 1.01$ and $OR = 1.30$), as well as random effects of participants and test items on intercepts. In the process of model selection, declarative memory, as a main effect or in interactions, was not found to improve the model's fit, leading to the selection of a more economical model. For an illustration of the extent to which estimates would vary in a model including DECL, see Appendix C.2 (1–3).

Overall, the model highlights that, when outcomes were controlled for pretest scores and time spent practicing on the platform, posttest scores were significantly positively predicted by an interaction between implicit statistical learning ability and WM. Specifically, the interaction can be interpreted to show that higher working memory was associated with better accuracy, but only for those learners with higher implicit statistical learning abilities (Fig. 2).

3.4. RQ2-RQ3b

The second model explored the relationships between cognitive variables and language outcomes in the written task. In this analysis the linear mixed model did not converge, which led to the adoption of a simpler linear regression model (Table 2b). The model accounted for about 22 % of the variance.

Although with very small effect sizes, the analysis found positive relationships between both Gains (partial correlation squared; $R^2 = 0.05$) and EFL proficiency scores ($R^2 = 0.01$) and task outcomes. Similar to the previous analysis, the model returned a positive significant Implicit Statistical Learning × WM interaction (Fig. 3; $R^2 = 0.01$; SL and WM conditional effects $R^2 = 0.01$ and $R^2 = 0.04$). This significant effect indicates that, while WM generally has a positive association with TT3 scores, the magnitude of this association was larger for individuals with better implicit statistical learning abilities. Also in this case, adding declarative memory, as a main effect or in interactions, did not improve the model's fit. For an illustration of the extent to which estimates would vary in a model including DECL, see Appendix C.2 (4–6).

4. Discussion

The first analysis (RQ1-RQ3a) explored to what extent declarative memory (a component of explicit aptitude), implicit statistical learning (a component of implicit aptitude) and WM, or the interactions between them, predicted EFL development as measured by posttest scores. It found that posttest scores were predicted by a positive interaction between implicit statistical learning ability and WM.

Generally, these results are consistent with previous studies with instructed adolescents that reported a pivotal role of L2 implicit aptitude (e.g., Frost et al., 2013; Kaufmann et al., 2010). Specifically, the interaction points to a multiplicative effect between WM and

Table 2a

Generalized Mixed-Effects Model of the Effects of Cognitive IDs, Pretest Scores, and Time (Digital Instruction) on Accuracy in the Posttest ($R^2\Delta = .36$; C.N. = 2.01; Cases N = 1460, Participants = 55).

Fixed effects	*β*	*SE*	*z*	95 % CI (Wald)		*p*
				lower	upper	
(Intercept)	−1.99	0.29	−6.86	−2.56	−1.42	<0.001***
Pretest	0.70	0.19	3.74	0.33	1.06	<0.001***
Time	0.48	0.18	2.67	0.13	0.84	0.008**
WM	0.26	0.20	1.33	−0.12	0.65	0.183
SL	0.01	0.17	0.08	−0.32	0.35	0.933
SL*WM	0.35	0.17	2.03	0.01	0.68	0.042*

Note. *$p < .05$; **$p < .01$; ***$p < .001$. Syntax: POST ACC ~ (1|PART) + (1|ITEMS) + PRETEST + TIME + SL * WM. C.N. = condition number. SL = implicit statistical learning ability; WM = working memory.

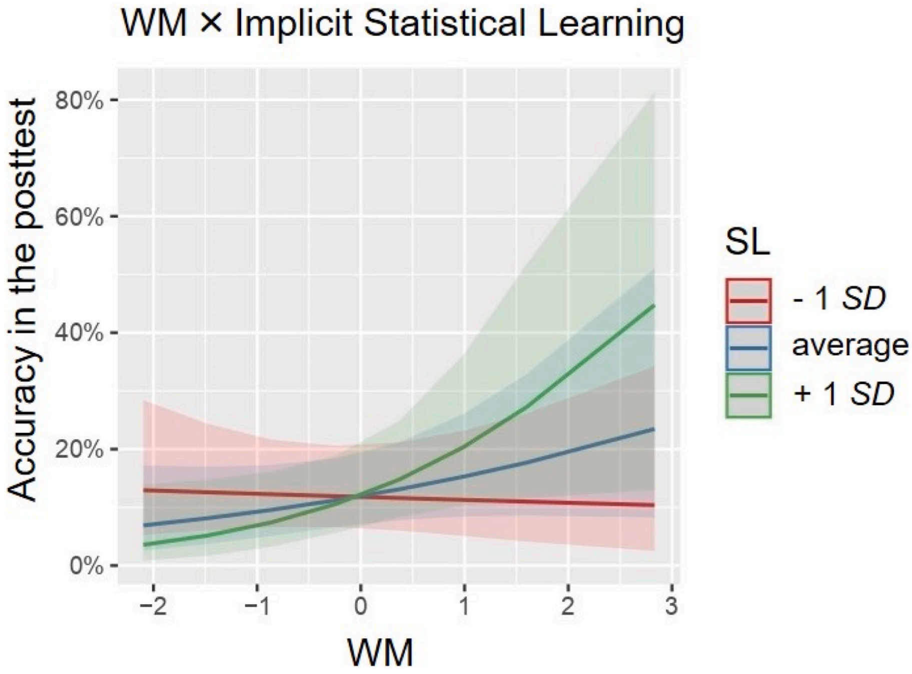

Fig. 2. Effect of the interaction between working memory and implicit statistical learning on the posttest outcomes.

Table 2b
Linear model of the effects of cognitive IDs, EFL proficiency and gains on task scores. (R^2 = 0.22; C.N. = 1.91; cases N = 1384, participants N = 55).

Fixed effects	*β*	*SE*	*z*	95 % CI (Wald)		*p*
				lower	upper	
(Intercept)	0.00	0.01	−0.03	−0.03	0.02	0.742
EFL Prof	−0.07	0.02	−4.21	−0.10	−0.04	<0.001***
Gains	0.15	0.01	9.58	0.12	0.17	<0.001***
WM	0.16	0.02	8.92	0.12	0.19	<0.001***
SL	0.07	0.01	4.71	0.04	0.10	<0.001***
SL*WM	0.07	0.01	4.60	0.04	0.10	<0.001***

Note. ****p* < .001. Syntax: QUESTIONS ~ EFL Prof + Gains + SL * WM. C.N. = condition number. SL = implicit statistical learning ability; WM = working memory.

implicit statistical learning in supporting L2 development, whereby learning was superior in those individuals with a large WM capacity who also had better implicit statistical learning abilities.

One way to interpret the interaction is to consider a possible role of implicit statistical learning and WM in the updating of linguistic representations of EFL direct questions. According to this interpretation, individuals with strong implicit statistical learning skills would not only have had access to more stable long-term implicit grammar representations, but would have also been more efficient at tracking the distributional properties of relevant input during the study's treatment period. Since the ability to update nontargetlike representations is key to L2 development, learners with larger WM capacity, in addition to strong implicit statistical learning, would have been particularly advantaged in their ability to retrieve, track, and update distributional information.

Note that a substantial role of implicit statistical learning (and in general of implicit aptitude) would also be expected considering that, independently of the extent to which L2 representations and use were targetlike, direct questions did not constitute new learning, as students would have had consistent and repeated exposure to exemplars of the construction during the previous four years of formal instruction (Grade 3 to Grade 6), in addition to the intensive input exposure in the practice phase. Although the immersive learning conditions found to play a decisive role in some previous studies (e.g., Bolibaugh & Foster, 2021; Faretta-Stutenberg & Morgan-Short, 2018) were not available in our case, the present results are consistent with the prediction of a significant relationship between measures of implicit aptitude and L2 development at higher levels of proficiency and input exposure (e.g., Hamrick et al., 2018; Morgan-Short et al., 2014). Unlike previous studies that analysed the early stages of exposure to a new L2 (e.g., Pili-Moss, 2022), rather than more advanced stages of learning, significant interactions between declarative memory and implicit aptitude components did not emerge.

A further result was that, contrary to our initial hypothesis, declarative memory did not significantly predict EFL development. This finding was unexpected in this group of learners in consideration of their age and of the conditions in which they were instructed and

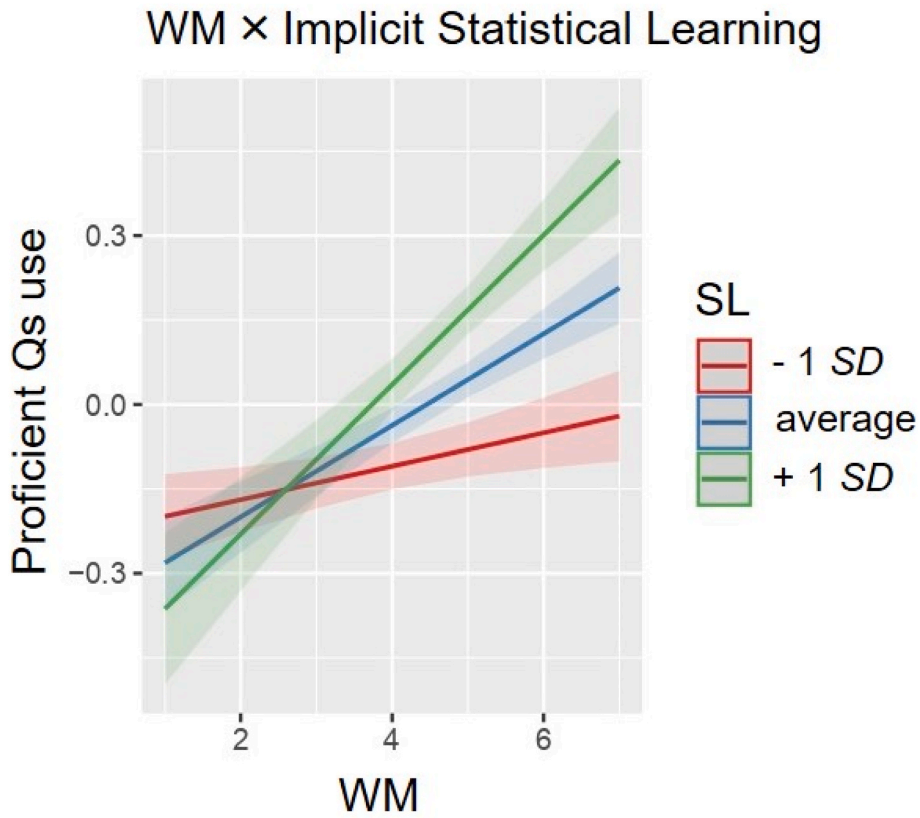

Fig. 3. Effect of the interaction between working memory and implicit statistical learning on the task outcomes.

tested. Previous research has found that the declarative memory abilities of early adolescents are more similar to those of young adults, rather than those of younger children (e.g., Thaler et al., 2013). Accordingly, one could have arguably expected adolescents to pattern with adults and more substantially rely on explicit learning strategies in instructed environments. Reliance on explicit learning skills should have also been reinforced by explicit instruction conditions in traditional frontal teaching (Leow, 2015), as well as in the digital practice, where the corrective feedback provided was clearly explicit and metalinguistic in nature. Finally, the completion of the digital written posttest was self-paced and untimed, which should also have facilitated the use of explicit learning strategies.

An alternative possibility that cannot be discarded is that other types of explicit aptitude (e.g., language analytic ability) played a more prominent role compared to declarative memory, per se. Although it was hypothesized that a measure of explicit sequence learning would be particularly suitable given the linguistic target (word order in questions), it is also possible that a different measure of declarative memory (e.g., one based on recollection rather than recognition) could have been more sensitive. Further, as suggested by a reviewer, the null results of the declarative learning task may be due to the mode of stimuli presentation in the exposure and the test phases, i.e., sequential rather than simultaneous, potentially inviting processing of transitional statistics despite the explicit learning conditions provided in the task (see e.g., Siegelman, Bogaerts, Christiansen, & Frost, 2017).

Turning to the covariates in the RQ1-RQ3a analysis, we found that proficiency in the pretest was a significant factor in determining posttest scores. Overall time spent practising on the platform was also a significant factor, thus directly positively relating aspects of engagement with the digital practice to L2 development.

The RQ2-RQ3b analysis explored the extent to which the investigated cognitive variables (or the interactions between them) predicted proficient use of EFL direct questions in a written communicative task, as measured by a composite index. With regard to the cognitive predictors, the findings returned a similar picture to the first analysis. Specifically, a significant, positive two-way Implicit Statistical Learning × WM interaction was obtained over and above the main effects of Implicit Statistical Learning and WM, with no evidence of a significant predictive role for declarative memory, or of an interaction between declarative memory and implicit statistical learning. Significant relationships between WM and task outcomes were previously found in description and narrative tasks (Mavrou, 2020; Vasylets & Marín, 2021). Further, at least one study (Granena, 2019) found that implicit aptitude predicted attainment in oral tasks, in conditions requiring real-time communicative L2 production which, albeit in a different modality, bear similarities to the ones implemented in the present study.

The Implicit Statistical Learning × WM interaction can be potentially accounted for in two ways. Along the lines of what proposed for RQ1-RQ3a, it can be explained as a synergetic relationship occurring during the instructional treatment that supports updating of long-term linguistic representations and subsequent task outcomes. However, it cannot be excluded that individuals with larger WM capacity were more efficient in retrieving and processing more stable implicit representations in real-time language production during a task performed under time pressure. A clearer elucidation of when exactly WM played a pivotal role, whether in the learning process, in task performance, or in both, cannot be fully resolved within the present study design and requires further investigation.

In contrast to some previous studies (e.g., Mujtaba, 2021; Yang et al., 2019) declarative memory was not found to significantly relate to proficiency in written production. Considerations relative to the age of the learners and the instruction conditions, that are similar to those discussed for the RQ1-RQ3a analysis, would apply again here. In addition, one could also suggest that in conditions of time pressure, such as those imposed by the communicative interaction, limited reliance on language knowledge gained and retrieved through declarative memory, and more reliance on implicit representations (with a direct role of implicit statistical learning in crystallizing them) might have occurred.

In the RQ2-RQ3b analysis, overall EFL proficiency at the end of the previous school year and gains at posttest were the two significant covariates included in the model providing the best fit, whereas pretest scores were not found to predict proficiency in the written task. The relevance of overall proficiency can be accounted for by considering that this measure would have better tapped a range of linguistic skills linked to the productive and communicative nature of the task, beyond grammar accuracy. The fact that gains at posttest, rather than pretest scores, were related to task outcomes, specifically links attainment in the task to learning that occurred during the study's instruction and practice period, rather than to knowledge available prior to treatment. This constitutes a pedagogically relevant finding because it suggests that L2 learning acquired in traditional classroom-based and largely non-communicative contexts, combined with intensive digital practice, can transfer to communicative classroom-based situations.

Overall, the findings relative to the posttest attainment and written task summarised in the present discussion point to a central role for implicit statistical learning ability in combination with WM capacity and to a lack of evidence that declarative memory significantly predicted L2 development or task outcomes. The fact that the interaction between implicit statistical learning ability and WM was a significant predictor, across both outcome measures, could indicate that the combination of these abilities was generally central to the process of L2 development in the learning and practice phase. Specifically, we suggested that both large WM capacity and high implicit statistical learning ability were crucial for encoding, retrieval, and/or updating of stable, implicit long-term representations of direct question constructions.

4.1. Limitations and further developments

The present study has a number of limitations that should be addressed in future research. Among the aspects deserving of attention, a first focus should be on the possibility to clearly discern between how cognitive individual differences impact L2 learning versus L2 use. To this end, future studies could include measures of task performance alongside measures of attainment in task outputs.

Although it was possible to establish a significant correlation between L2 outcomes and time spent on digital instruction in learners otherwise comparable in terms of the classroom instruction they had received, a control group who did not receive digital instruction was not available in the present study. Future research should further clarify the distinct contributions of class-instruction and digital instruction to L2 development. Similarly, manipulating corrective feedback provision during digital practice could shed light on the interactions between cognitive variables and corrective feedback effectiveness.

Turning to cognitive individual differences, a further limitation concerns the relatively low internal consistency of the implicit statistical learning task, an issue that has been discussed in the literature in relation to this type of task (e.g., Siegelman, Bogaerts, Christiansen, & Frost, 2017). Although values of Cronbach's alpha higher than 0.80 were reported in an adult study for the task we used as a model (Siegelman, Bogaerts, & Frost, 2017), these were not replicated. One possibility is that the participants' age group plays a role in the extent to which SL task reliability values are replicable (Arnon, 2020).

Beside the need to further investigate which components of explicit aptitude predict L2 development, a more analytic approach seems warranted also in regard to declarative memory. Future studies should include batteries gauging a wider range of aspects of declarative memory, including both episodic (e.g., recognition and recollection of associations in different modalities) and semantic memory, since they appear to make different contributions to L2 development (Murphy et al., 2021). Relatedly, an aspect that should also be explored to further theoretical development in the field of language aptitude studies is the extent to which a range of components of implicit aptitude converge in predicting L2 development in studies with a design similar to the present one.

4.2. Conclusion

Aim of the present study was to investigate the relationship between measures of EFL attainment (posttest gains and L2 use in a written task) and cognitive language aptitude, including memory-related aspects of explicit aptitude, implicit aptitude and WM, in L1 German adolescents administered a combination of regular classroom instruction and digital practice.

The core findings indicated that, independently of the outcome measure considered, a positive interaction between implicit statistical learning and WM significantly predicted language outcomes. The results demonstrate that, despite the importance of explicit learning in the classroom (Leow, 2015), implicit aptitude and WM are relevant individual differences in mainstream instructed learning beyond immersion contexts. They also underscore the need to systematically explore interactions among cognitive variables if we wish to gain a clearer understanding of processes underpinning cognitive language aptitude in L2 learning.

CRediT authorship contribution statement

Diana Pili-Moss: Writing – review & editing, Writing – original draft, Visualization, Validation, Software, Project administration,

Methodology, Investigation, Formal analysis, Data curation, Conceptualization. **Phillip Hamrick:** Writing – review & editing, Conceptualization. **Katharina Wendebourg:** Writing – review & editing. **Torben Schmidt:** Writing – review & editing, Resources, Project administration, Methodology, Funding acquisition. **Detmar Meurers:** Software, Project administration, Methodology, Funding acquisition.

Funding

This study was funded by the German Federal Ministry of Education and Research (BMBF) - grant numbers 01JD1905A and 01JD1905B.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.system.2025.103656.

References

An, H., & Li, S. (2022). Working memory in second language interaction. In J. W. Schwieter, & Z. Edward Wen (Eds.), *The cambridge handbook of working memory and language* (pp. 656–697). Cambridge: Cambridge University Press. https://doi.org/10.1017/9781108955638.036.

Arnon, I. (2020). Do current statistical learning tasks capture stable individual differences in children? An investigation of task reliability across modality. *Behavior Research Methods, 52*, 68–81. https://doi.org/10.3758/s13428-019-01205-5

Baddeley, A. (2012). Working memory: Theories, models, and controversies. *Annual Review of Psychology, 63*, 1–29. https://doi.org/10.1146/annurev-psych-120710-100422

Bates, D., Mächler, M., Bolker, B., & Walker, S. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software, 67*(1), 1–48. https://doi.org/10.18637/jss.v067.i01

Bolibaugh, C., & Foster, P. (2021). Implicit statistical learning in naturalistic and instructed morphosyntactic attainment: An aptitude-treatment interaction design. *Language Learning, 71*, 959–1003. https://doi.org/10.1111/lang.12465

Brill-Schuetz, K. A., & Morgan-Short, K. (2014). The role of procedural memory in adult second language acquisition. In P. Bello, M. Guarini, M. McShane, & B. Scassellati (Eds.), *Proceedings of the 36th annual conference of the cognitive science society* (pp. 260–265). Austin: Texas: Cognitive Science Society.

Brooks, P. J., Kwoka, N., & Kempe, V. (2017). Distributional effects and individual differences in L2 morphology learning. *Language Learning, 67*, 171–207. https://doi.org/10.1111/lang.12204

Carroll, J., & Sapon, S. (2002). *Manual for the MLAT*. Second Language Testing, Inc.

Cotton, K., & Ricker, T. J. (2021). Working memory consolidation improves long-term memory recognition. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 47*(2), 208–219. https://doi.org/10.1037/xlm0000954

Cowan, N. (2008). What are the differences between long-term, short-term, and working memory? *Progress in Brain Research, 169*, 323–338. https://doi.org/10.1016/S0079-6123(07)00020-9

Ellis, N. (2015). Implicit and explicit language learning: Their dynamic interface and complexity. In P. Rebuschat (Ed.), *Implicit and explicit learning of languages* (pp. 3–23). John Benjamins. https://doi.org/10.1075/sibil.48.01ell.

Ellis, R. (2021). Options in a task-based language-teaching curriculum: An educational perspective. *TASK, 1*(1), 11–46. https://doi.org/10.1075/task.00002.ell

Faretta-Stutenberg, M., & Morgan-Short, K. (2018). The interplay of individual differences and context of learning in behavioral and neurocognitive second language development. *Second Language Research, 34*(1), 67–101. https://doi.org/10.1177/0267658316684903

Frost, R., Siegelman, N., Narkiss, A., & Afek, L. (2013). What predicts successful literacy acquisition in a second language? *Psychological Science, 24*(7), 1243–1252. https://doi.org/10.1177/0956797612472207

Granena, G. (2019). Cognitive aptitudes and L2 speaking proficiency: Links between LLAMA and Hi-LAB. *Studies in Second Language Acquisition, 41*, 313–336. https://doi.org/10.1017/S0272263118000256

Hamrick, P. (2015). Declarative and procedural memory abilities as individual differences in incidental language learning. *Learning and Individual Differences, 44*, 9–15. https://doi.org/10.1016/j.lindif.2015.10.003

Hamrick, P., Lum, J. A. G., & Ullman, M. T. (2018). Child first language and adult second language are both tied to general-purpose learning systems. *Proceedings of the National Academy of Sciences, 115*(7), 1487–1492. https://doi.org/10.1073/pnas.1713975115

Hui, B., Rudzewitz, B., & Meurers, D. (2023). Learning processes in interactive CALL systems: Linking automatic feedback, system logs, and learning outcomes. *Language, Learning and Technology, 27*(1), 1–23. https://hdl.handle.net/10125/73527.

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

Kenanidis, P., Llompart, M., Pili-Moss, D., & Dabrowska, E. (2024). *Implicit-statistical and explicit learning as predictors of early L2 acquisition* [Poster session]. *Interdisciplinary advances in statistical learning*. San Sebastián, Spain.

Kersten, K., & Winsler, A. (2023). *Understanding variability in second language acquisition, bilingualism, and cognition: A multi-layered perspective*. Routledge.

Kim, M., Tian, Y., & Crossley, S. A. (2021). Exploring the relationships among cognitive and linguistic resources, writing processes, and written products in second language writing. *Journal of Second Language Writing, 53*, Article 100824. https://doi.org/10.1016/j.jslw.2021.100824

Kormos, J. (2012). The role of individual differences in L2 writing. *Journal of Second Language Writing, 21*, 390–403. https://doi.org/10.1016/j.jslw.2012.09.003

Kormos, J., & Sáfár, A. (2008). Phonological short-term memory, working memory and foreign language performance in intensive language learning. *Bilingualism: Language and Cognition, 11*, 261–271. https://doi.org/10.1017/S1366728908003416

Kormos, J., & Trebits, A. (2012). The role of task complexity, modality, and aptitude in narrative task performance. *Language Learning, 62*, 439–472. https://doi.org/10.1111/j.1467-9922.2012.00695.x

Leow, R. P. (2015). *Explicit learning in the L2 classroom: A student-centered approach* (1st ed.). Routledge. https://doi.org/10.4324/9781315887074

Li, S. (2015). The differential roles of language analytic ability and working memory in mediating the effects of two types of feedback on the acquisition of an opaque linguistic structure. In C. Sanz, & B. Lado (Eds.), *Individual differences, L2 development & language program administration: From theory to application* (pp. 32–52). Cengage Learning.

Li, S. (2016). The construct validity of language aptitude: A meta-analysis. *Studies in Second Language Acquisition, 38*(4), 801–842. https://doi.org/10.1017/S027226311500042X

Li, S. (2017). The effects of cognitive aptitudes on the process and product of L2 interaction: A synthetic review. In L. Gurzynski-Weiss (Ed.), *Expanding individual difference research in the interaction approach: Investigating learners, instructors, and other interlocutors* (pp. 41–70). John Benjamins. https://doi.org/10.1075/aals.16.03li.

Li, S. (2022). Working memory and second language learning: A critical and synthetic review. In A. Godfroid, & H. Hopp (Eds.), *The routledge handbook of second language acquisition and psycholinguistics* (pp. 348–360). Taylor & Francis. 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*(3), 473–497. https://doi.org/10.1017/S0272263121000024

Linck, J., Hughes, M., Campbell, S., Silbert, N., Tare, M., Jackson, S., et al. (2013). Hi-LAB: A new measure of aptitude for high-level language proficiency. *Language Learning, 63*, 530–566. https://doi.org/10.1111/lang.12011

Linck, J. A., Osthus, P., Koeth, J. T., & Bunting, M. F. (2014). Working memory and second language comprehension and production: A meta-analysis. *Psychonomic Bulletin & Review, 21*(4), 861–883. https://doi.org/10.3758/s13423-013-0565-2

Manchón, R. M., McBride, S., Mellado Martínez, M. D., & Vasylets, O. (2023). Working memory, L2 proficiency, and task complexity: Independent and interactive effects on L2 written performance. *Studies in Second Language Acquisition, 45*(3), 737–764. https://doi.org/10.1017/S0272263123000141

Massonnié, J., Rogers, C. J., Mareschal, D., & Kirkham, N. Z. (2019). Is classroom noise always bad for children? The contribution of age and selective attention to creative performance in noise. *Frontiers in Psychology, 10*, Article 381. https://doi.org/10.3389/fpsyg.2019.00381

Mavrou, I. (2020). Working memory, executive functions, and emotional intelligence in second language writing. *Journal of Second Language Writing, 50*, Article 100758. https://doi.org/10.1016/j.jslw.2020.100758

McDonough, K., & Trofimovich, P. (2016). The role of statistical learning and working memory in L2 speakers' pattern learning. *The Modern Language Journal, 100*(2), 428–445. http://www.jstor.org/stable/44135020.

Meara, P. M., & Rogers, V. E. (2019). *The LLAMA tests v3*. Lognostics. https://www.lognostics.co.uk/tools/LLAMA_3/index.htm.

Meurers, D., Kuthy, K. D., Nuxoll, F., Rudzewitz, B., & Ziai, R. (2019). Scaling up intervention studies to investigate real-life foreign language learning in school. *Annual Review of Applied Linguistics, 39*, 161–188. https://doi.org/10.1017/S0267190519000126

Misyak, J. B., & Christiansen, M. H. (2012). Statistical learning and language: An individual differences study. *Language Learning, 62*, 302–331. https://doi.org/10.1111/j.1467-9922.2010.00626.x

Morgan-Short, K., Faretta-Stutenberg, M., Brill-Schuetz, K. A., Carpenter, H., & Wong, P. C. M. (2014). Declarative and procedural memory as individual differences in second language acquisition. *Bilingualism: Language and Cognition, 17*, 56–72. https://doi.org/10.1017/S1366728912000715

Mujtaba, S. M., Kamyabi Gol, A., & Parkash, R. (2021). A study on the relationship between language aptitude, vocabulary size, working memory, and L2 writing accuracy. *Foreign Language Annals, 54*, 1059–1081. https://doi-org.ezproxy.lancs.ac.uk/10.1111/flan.12584.

Murphy, J., Miller, R. T., & Hamrick, P. (2021). Contributions of declarative memory and prior knowledge to incidental vocabulary learning. *The Mental Lexicon, 16*(1), 49–68. https://doi.org/10.1075/ml.20012.mur

Park, J., Yoon, H.-D., Yoo, T., Shin, M., & Jeon, H.-A. (2020). Potential and efficiency of statistical learning closely intertwined with individuals' executive functions: A mathematical modeling study. *Scientific Reports, 10*, Article 18843. https://doi.org/10.1038/s41598-020-75157-8

Pedraza, F., et al. (2024). Evidence for a competitive relationship between executive functions and statistical learning. *Npj Science of Learning, 9*, 30. https://doi.org/10.1038/s41539-024-00243-9

Peng, A., Orosco, M. J., Wang, H., Swanson, H. L., & Reed, D. K. (2021). Cognition and writing development in early adolescent English learners. *Journal of Educational Psychology, 114*, 1136–1155. https://doi.org/10.1037/edu0000695

Pili-Moss, D. (2021). Cognitive predictors of child second language comprehension and syntactic learning. *Language Learning, 71*(3), 907–945. https://doi.org/10.1111/lang.12454

Pili-Moss, D. (2022). Long-term memory predictors of adult language learning at the interface between syntactic form and meaning. *PLoS One, 17*(10), Article e0275061. https://doi.org/10.1371/journal.pone.0275061

Pili-Moss, D., Brill-Schuetz, K. A., Faretta-Stutenberg, M., & Morgan-Short, K. (2020). Contributions of declarative and procedural memory to accuracy and automatization during second language practice. *Bilingualism: Language and Cognition, 23*(3), 639–651. https://doi.org/10.1017/S1366728919000543

Pili-Moss, D., Schmidt, T., & Beilharz, S. (2023). Projekt Interact for School (I4S). Unterrichtseinheit und Target-Task zum Thema „ Feelings and Friendship. *Language Focus: Question Formation; Grade 7. Leuphana Universität Lüneburg/Universität Tübingen*. https://doi.org/10.48548/pubdata-109

Plonsky, L., & Ghanbar, H. (2018). Multiple regression in L2 research: A methodological synthesis and guide to interpreting R2 values. *The Modern Language Journal, 102*(4), 713–731. https://doi.org/10.1111/modl.12509

Poldrack, R. A., & Packard, M. G. (2003). Competition among multiple memory systems: Converging evidence from animal and human brain studies. *Neuropsychologia, 41*, 245–251. https://doi.org/10.1016/S0028-3932(02)00157-4

R Core Team. (2021). *R: A language and environment for statistical computing*. Vienna, Austria: R Foundation for Statistical Computing. https://www.R-project.org/.

Ren, J., & Wang, M. (2023). Development of statistical learning ability across modalities, domains, and languages. *Journal of Experimental Child Psychology, 226*. https://doi.org/10.1016/j.jecp.2022.105570

Révész, A., Michel, M., & Lee, M. (2023). Exploring the relationship of working memory to the temporal distribution of pausing and revision behaviors during L2 writing. *Studies in Second Language Acquisition, 45*, 680–709. https://doi.org/10.1017/S0272263123000074

Roehr-Brackin, K., & Tellier, A. (2019). The role of language-analytic ability in children's instructed second language learning. *Studies in Second Language Acquisition, 41*(5), 1111–1131. https://doi.org/10.1017/S0272263119000214

Schmidt, T., & Strasser, T. (2022). Artificial intelligence in foreign language learning and teaching: a CALL for intelligent practice. *Anglistik: International Journal of English Studies, 33*(1), 165–184. https://doi.org/10.33675/ANGL/2022/1/14

Schwieter, J. W., & Wen, Z. (2022). *The cambridge handbook of working memory and language*. Cambridge University Press. https://doi.org/10.1017/9781108955638

Siegelman, N., Bogaerts, L., Christiansen, M. H., & Frost, R. (2017). Towards a theory of individual differences in statistical learning. *Philosophical Transactions of the Royal Society B, 372*, Article 20160059. https://doi.org/10.1098/rstb.2016.0059

Siegelman, N., Bogaerts, L., & Frost, R. (2017). Measuring individual differences in statistical learning: Current pitfalls and possible solutions. *Behavior Research Methods, 49*(2), 418–432. https://doi.org/10.3758/s13428-016-0719-z

Skehan, P. (2002). Theorising and updating aptitude. In P. Robinson (Ed.), *Individual differences in instructed language learning* (pp. 69–95). John Benjamins. https://doi.org/10.1075/lllt.2.06ske.

Tagarelli, K. M., Mota, M. B., & Rebuschat, P. (2011). The role of working memory in implicit and explicit language learning. *Proceedings of the Annual Meeting of the Cognitive Science Society, 33*. https://escholarship.org/uc/item/0r55c3fk.

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

Thaler, N. S., Allen, D. N., Reynolds, C. R., & Mayfield, J. (2013). Identifying neurodevelopmental stages of memory from childhood through adolescence with cluster analysis. In D. Allen, & G. Goldstein (Eds.), *Cluster analysis in neuropsychological research*. Springer. https://doi.org/10.1007/978-1-4614-6744-1_4.

Ullman, M. T. (2020). The declarative/procedural model. In B. VanPatten, G. D. Keating, & S. Wulff (Eds.), *Theories in second language acquisition* (3rd ed., pp. 128–161). Abingdon: Routledge. https://doi.org/10.4324/9780429503986-7.

Vasileiou, I., & Pili-Moss, D. (2022). The role of learners' memory in app-based language instruction: The case of Duolingo. In B. Arnbjörnsdóttir, B. Bédi, L. Bradley, K. Friðriksdóttir, H. Garðarsdóttir, S. Thouësny, et al. (Eds.), *Intelligent CALL, granular systems, and learner data: Short papers from EUROCALL 2022* (pp. 364–369). https://doi.org/10.14705/rpnet.2022.61.1485

Vasylets, O., & Marín, J. (2021). The effects of working memory and L2 proficiency on L2 writing. *Journal of Second Language Writing, 52*, Article 100786. https://doi.org/10.1016/j.jslw.2020.100786

Vasylets, O., Mellado, M. D., & Plonsky, L. (2022). The role of cognitive individual differences in digital versus pen-and-paper writing. *Studies in Second Language Learning and Teaching, 12*, 723–745. https://doi.org/10.14746/ssllt.2022.12.4.9

Williams, J. N. (2015). Working memory in SLA research: Challenges and prospects. In Z. Wen, M. Borges Mota, & A. Mc Neill (Eds.), *Working memory in second language acquisition and processing* (pp. 301–308). Multilingual Matters. https://doi.org/10.21832/9781783093595.

Yang, Y., Sun, Y., Chang, P., & Li, Y. (2019). Exploring the relationship between language aptitude, vocabulary size, and EFL graduate students' L2 writing performance. *Tesol Quarterly, 53*, 845–856. https://doi.org/10.1002/tesq.510
Yi, B., & Ni, C. (2015). Planning and working memory effects on L2 performance in Chinese EFL learners' argumentative writing. *Indonesian Journal of Applied Linguistics, 5*, 44–53.
Zabihi, R. (2018). The role of cognitive and affective factors in measures of L2 writing. *Written Communication, 35*, 32–57. https://doi.org/10.1177/0741088317735836
Zalbidea, J. (2017). "One task fits all?" The roles of task complexity, modality, and working memory capacity in L2 performance. *The Modern Language Journal, 101*, 335–352. https://doi.org/10.1111/modl.12389